МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА

А. Є. АЧКАСОВ
О. О. ВОРОНКОВ

КОНСПЕКТ ЛЕКЦІЙ
з курсу

ВИЩА ТА ПРИКЛАДНА МАТЕМАТИКА (ДОСЛІДЖЕННЯ ОПЕРАЦІЙ 1)

(для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030601 – Менеджмент)

Харків
ХНУМГ ім. О. М. Бекетова
2015

Воронков О. О. Конспект лекцій з курсу «Вища та прикладна математика (Дослідження операцій 1)» (для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030601 – Менеджмент) / О. О. Воронков, А. Є. Ачкасов; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. - Харків: ХНУМГ, 2015. - 110 с.

Автори: канд. екон. наук, доц. О. О. Воронков
д-р екон. наук, проф. А. Є. Ачкасов

Рецензент: канд. екон. наук, доцент Г. І. Базецька

Рекомендовано кафедрою економіки підприємств міського господарства, протокол № 1 від 27.08.2014 р.

ЗМІСТ

ВСТУП

Курс «Вища та прикладна математика (Дослідження операцій 1)» є нормативною дисципліною циклу природничо-наукової і загальноекономічної підготовки в навчальному плані за напрямом 6.030601 «Менеджмент» для кваліфікаційного рівня «Бакалавр». Обсяг курсу становить 72 академічних години або 2 кредити ESTC. При вивченні за заочною формою обсяг аудиторних занять становить 8 годин (4 години лекцій і 4 години практичних занять), на самостійну роботу студента припадає 64 години. Підсумковий контроль знань (екзамен) проводиться в письмовій формі. У процесі вивчення курсу студенти повинні виконати контрольну роботу.

Метою вивчення дисципліни «Вища та прикладна математика (Дослідження операцій 1)» є формування системи базових знань в області методології постановки економічних задач та побудови математичних моделей і методів їх розв'язання й аналізу.

В результаті вивчення курсу студенти повинні оволодіти прийомами побудови економіко-математичних моделей, основними математичними поняттями і методами розв'язання оптимізаційних задач різної складності, зокрема наступними: задачі управління запасами, задачі масового обслуговування, задачі упорядкування та координації, задачі та моделі заміни, імітаційне моделювання, прийняття рішень в умовах невизначеності та ризику, задачі теорії ігор, проблеми багатокритеріальної оптимізації.

Сучасна економічна наука характеризується широким використанням математики. Математичні методи стали складовою частиною методів будь-якої економічної науки, включаючи економічну теорію. Використання математичних методів відчиняє нові можливості, і фахівцеві необхідно вміти формулювати та розв'язувати задачі з оптимізації виробництва, моделювати економічну динаміку, статистично оцінювати економічні залежності, а також користуватися ігровими методами.

Витоки дослідження операцій йдуть у далеку історію. Різке збільшення обсягів виробництва, розподіл праці зумовив поступову диференціацію управлінської праці. З'явилася необхідність в плануванні матеріальних, трудових та грошових ресурсів, в обліку та аналізі праці і виробленні прогнозу на майбутнє. В управлінському апараті почали виділятися підрозділи: відділ фінансів, збуту, бухгалтерії, планово-економічний відділ та інші, що прийняли на себе окремі управлінські функції.

До цього періоду належать перші роботи з досліджень у сфері організації праці та управління. Як самостійний науковий напрям дослідження операцій оформилося на початку 40-х років. Перші публікації з дослідження операцій належать 1939-1940 рокам, в яких ці методи застосовані для вирішення військових завдань, зокрема для аналізу й дослідження військових операцій. Звідси і пішла назва дисципліни. Пізніше принципи і методи дослідження операцій стали застосовуватися у сфері промислово-фінансового управління. Із збільшенням масштабів виробництва розширювалися масштаби операційних досліджень, коло вирішуваних задач, удосконалювалися методи нової науки. Виникла необхідність в підготовці фахівців з дослідження операцій - операціоністів. У провідних університетах США та Англії вперше було почато систематичне викладання курсу дослідження операцій. Виникла необхідність в координації роботи операціоністів, у регулярному обміні теоретичними дослідженнями та прикладними розробками. З цією метою в 1957 р. було створено Міжнародну федерацію дослідження операцій IFORS, до складу якої входили національні товариства і комітети з дослідження операцій багатьох країн.

Нерідко дослідження операцій розуміють як застосування математичних кількісних методів для обґрунтування рішень в усіх областях людської діяльності. При цьому випускають з уваги, що дослідження операцій як наукова дисципліна знаходиться на стику наук, що оперують не тільки кількісними, але й якісними факторами. Саме зневагою до якісних факторів пояснюється основна кількість негативних результатів застосування методів дослідження операцій до розв’язання практичних задач, оскільки змістові і методологічні аспекти в будь-якому операційному дослідженні грають анітрохи не меншу роль за формальні.

ЗМ 1 Детерміновані моделі і методи

ТЕМА 1
ПРЕДМЕТ ТА ЗАДАЧІ ДОСЛІДЖЕННЯ ОПЕРАЦІЙ

Дослідження операцій є наукою, що займається розробкою і практичним застосуванням методів оптимального управління організаційними системами. В даний час методи дослідження операцій знаходять широке застосування у розв'язанні самих різних практичних задач, починаючи від перспективного планування наукових розробок та закінчуючи прогнозуванням розвитку сфери обслуговування. Це пов'язано з тим, що будь-яка операція є сукупністю цілеспрямованих дій, а дослідження операцій - пошук шляхів досягнення однієї або кількох цілей.

Предметом дослідження операцій є системи організаційного управління або організації, які складаються з великої кількості підрозділів, що взаємодіють один з одним, та інтереси яких не завжди узгоджуються і можуть бути навіть протилежними.

Мета дослідження операцій - кількісне обґрунтування ухвалюваних рішень з управління організаціями. Рішення, яке виявляється найвигіднішим для всієї організації, називають оптимальним, а рішення найвигідніше одному або кільком підрозділам називають субоптимальним.

Відзначимо основні особливості дослідження операцій. По-перше - це системний підхід до аналізу поставленої проблеми. Системний аналіз є основним методологічним принципом дослідження операцій, який полягає в тому, що будь-яке завдання, яким би частковим воно не здавалося, розглядається з погляду його впливу на критерій ефективності функціонування всієї системи. Другою характерною особливістю дослідження операцій є те, що під час розв'язання кожної проблеми виникають все нові і нові завдання. Якщо спочатку покладають вузькі цілі, застосування операційних методів є неефективним. Найбільшого ефекту можна досягнути тільки при безперервному дослідженні, що забезпечує спадкоємність в переході від одного завдання до іншого.

Дослідження операцій – це напрям в дослідженні і проектуванні систем, заснований на математичному моделюванні процесів і явищ. Дослідження операцій передбачає системний підхід, що полягає в пошуку існуючих взаємодій при оцінці діяльності або стратегії будь-якої частини організації. Висновки дослідження операцій в застосуванні до конкретних систем даються на підставі математичних моделей систем. При побудові моделей прагнуть виразити критерій, що характеризує якість функціонування системи, через керовані і некеровані (залежні і незалежні) змінні. З урахуванням реально діючих обмежень на змінні багато завдань дослідження операцій збігаються до завдань математичного програмування. У дослідженні операцій використовуються так само методи імітаційного моделювання складних систем, теорії масового обслуговування, теорії ігор і статистичних рішень, математичної статистики та ін. Розрізняють такі завдання дослідження операцій: завдання розподілу, управління запасами, заміни, масового обслуговування, впорядкування і координації, вибору маршруту, завдання пошуку, змагальні завдання та ін. Окрім загальних методів в дослідженні операцій розвиваються методи, що відповідають різним областям застосування (транспорт, постачання, торгівля та ін.).

Істотною особливістю дослідження операцій є прагнення знайти оптимальне рішення поставленої задачі. Проте, часто таке рішення опиняється недосяжним із-за обмежень, що накладаються наявними ресурсами або рівнем сучасної науки. Наприклад, для комбінаторних завдань, зокрема завдань з календарного планування при кількості верстатів понад 4, оптимальне рішення при сучасному рівні розвитку математики можливо знайти лише простим перебором варіантів. Проте навіть при невеликих n кількість можливих варіантів настільки велика, що перебір всіх варіантів при існуючих обмеженнях на швидкодію ЕОМ та припустимий машинний час практично немислимі. Тоді доводиться обмежуватися пошуком досить хорошого або субоптимального рішення.

Особливість операційних досліджень полягає ще в тому, що їх проводять комплексно, за багатьма напрямами. Для проведення такого дослідження створюють операційну групу, у склад якої входять фахівці різних областей: інженери, математики, економісти, соціологи, психологи.

1.1 Основні поняття. Моделі дослідження операцій. Етапи дослідження операцій

Нагадаємо, що будь-яка математична модель є описом (часто наближеним) будь-якого класу явищ реального світу, вираженим за допомогою математичних символів.

Дослідження операцій - комплексна математична дисципліна, що займається побудовою, аналізом та застосуванням математичних моделей ухвалення оптимальних (якнайкращих в якомусь сенсі) рішень. Основним завданням дослідження операцій є завдання вибору в заданій множині елементу, що задовольняє тим або іншим критеріям (від грецького kriterion - засіб для порівняння). При цьому будь-який елемент множини називають припустимим рішенням, а вибраний елемент - оптимальним рішенням.

Побудова математичної моделі будь-якої задачі дослідження операцій завжди починається з опису множини D припустимих розв’язків та критеріїв оптимальності. При цьому під критерієм оптимальності розуміють ознаку, на підставі якої проводиться порівняльна оцінка припустимих розв’язків і вибір оптимального розв’язку.

На наступному етапі досліджень з використанням побудованої моделі треба знайти оптимальний розв’язок з множини D припустимих розв’язків. Але, як правило, ухвалює рішення не дослідник операції, який лише готує інформацію для ухвалення рішення а «особа, що приймає рішення» (ОПР). Саме ОПР формулює вимоги і вона ж ухвалює остаточне рішення.

Під ОПР розуміють групу людей, яка виробляє узгоджені вимоги до припустимих рішень та критеріїв оптимальності, ухвалює рішення. При цьому потрібно пам'ятати про те, що будь-яке рішення завжди ухвалюється відповідно до інформаційного стану ОПР, тобто відповідно до його змістових уявлень про можливі і доцільні дії в певних умовах. У математичній моделі це відображається на множині D припустимих рішень і критеріях оптимальності.

Треба зазначити, що побудова будь-якої математичної моделі задачі дослідження операцій майже завжди пов'язана з необхідністю задоволення двох, по суті, суперечливих вимог:

а) якомога точніше відобразити реальні явища, що вивчаються;

б) побудувати досить просту математичну модель, що дозволяє вирішити початкове завдання і одержати осяжні результати.

Саме тому на етапі побудови математичної моделі необхідна тісна співпраця дослідника операції (математика) і ОПР (фахівця).

Охарактеризуємо основні етапи операційного дослідження.

Першим етапом є постановка завдання. Спочатку завдання формулюють з погляду замовника. Під час аналізу системи завдання поступово уточнюється.

Другий етап - формалізація завдання. Одержавши доволі строгу і логічно несуперечливу змістовну постановку завдання, потрібно побудувати її математичну модель.

Третім етапом є знаходження методу рішення. Для знаходження оптимального рішення залежно від структури завдання застосовують ті або інші методи теорії оптимальних рішень, звані також методами математичного програмування.

Четвертий етап полягає у перевірці і коректуванні моделі. У складних системах, до яких належать системи організаційного типа, модель лише частково відображає реальний процес. Тому необхідна перевірка ступеня відповідності або адекватності моделі і реального процесу. Перевірку проводять шляхом порівняння передбаченої поведінки з фактичною при зміні значень зовнішніх некерованих дій. Коректування може зажадати додаткових досліджень об'єкту, уточнення структури математичної моделі, чисельних змін змінних. Отже, 4 етапи багато разів повторюються, поки не буде досягнуто задовільної відповідності між виходами об'єкту і моделі.

П’ятий етап - реалізація знайденого рішення на практиці. Впровадження можна розглядати як самостійне завдання, застосувавши до нього системний підхід і аналіз.

1.2 Класифікація задач та методів дослідження операцій

Зупинимося на класифікації можливих постановок завдань дослідження операцій, яка може проводитися за різними ознаками.

Класифікація завдань дослідження операцій за видом інформаційного стану ОПР. Якщо прийняття рішення відбувається в наперед відомому інформаційному стані ОПР, що не змінюється в часі, то завдання дослідження операцій називають статичним, а всю процедуру прийняття рішення можна реалізувати за один етап (крок). До класу статичних завдань належать завдання з кількома різними, але не змінними у часі інформаційними станами, це характерно для групи ОПР.

Якщо у процесі прийняття рішення інформаційний стан ОПР змінюється в часі, то завдання дослідження операцій називають динамічним. У цьому випадку найбільш доцільною є поетапна (багатокрокова) процедура прийняття рішення.

Класифікація завдань дослідження операцій за структурою інформаційного стану ОПР. Інформаційний стан ОПР може відповідати або єдиному фізичному стану об'єкта досліджень, або множині фізичних станів. У першому випадку завдання дослідження операцій називають детермінованим, в другому - стохастичним (або завданням ухвалення рішень в умовах ризику), якщо відомі апріорні імовірності знаходження об'єкту досліджень в кожному із станів, і невизначеним (або завданням ухвалення рішень в умовах невизначеності), якщо відсутня інформація про апріорну імовірність. Завдання ухвалення рішень в умовах невизначеності є предметом дослідження теорії ігор.

Отже, детермінований рівень - найпростіший рівень інформації про ситуацію, коли умови, в яких ухвалюються рішення, відомі повністю. Стохастичний рівень - рівень, при якому відома множина можливих варіантів умов та їх імовірнісний розподіл. Невизначений рівень - рівень, коли відома множина можливих варіантів, але без будь-якої інформації про їх імовірність.

Класифікація завдань дослідження операцій за видом критерію оптимальності. Класифікація завдань дослідження операцій може бути пов'язаною з видом використовуваного критерію оптимальності, який може мати будь-який вигляд, у тому числі і вигляд, що не формалізується.

Критерієм оптимальності може бути вимога про максимізацію або мінімізацію певної скалярної функції *f*, визначеної на множині припустимих рішень, її називають цільовою функцією. В цьому випадку завдання дослідження операцій називають завданням математичного програмування. Якщо ж критерієм оптимальності є вимога про максимізацію або мінімізацію кількох скалярних функцій, то говорять про завдання багатокритеріальної (векторної) оптимізації.

У математичному програмуванні частіше за інші розглядають завдання, в яких множина D припустимих рішень є підмножиною R^n, що задовольняє системі лінійних нерівностей

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = \overline{1,m}..$$

Таку множину, якщо вона непорожня і обмежена, називають опуклим багатогранником, оскільки цей клас опуклих множин відповідає геометричному уявленню про багатогранники в просторі.

Якщо множина припустимих рішень $D \in R^n$ є опуклим багатогранником, а цільова функція *f* лінійна, то початкове завдання називають завданням лінійного програмування. Якщо множина $D \in R^n$ є опуклим багатогранником, а цільова функція f є квадратичною, то початкове завдання називають завданням квадратичного програмування. Якщо $D \in R^n$ - опукла множина, а f - опукла функція, то початкове завдання називають завданням опуклого програмування. Теорія рішення стохастичних задач лінійного програмування є предметом досліджень стохастичного програмування.

У інших завданнях математичного програмування множина D припустимих рішень може бути кінцевою множиною. Такі завдання належать до дискретного програмування. У них припустимі рішення можуть бути точками цілочислових решіток Z^n (цілочислове програмування) або векторами, кожна координата яких може приймати лише два значення (булеве програмування). У окремих завданнях елементи множини D припустимих рішень можуть бути перестановками кінцевої кількості символів та ін.

Множина D припустимих рішень може бути підмножиною певного функціонального простору. В цьому випадку маємо завдання варіаційного числення або завдання оптимального управління.

Особливим випадком завдань математичного програмування є задачі про знаходження максиміну

$$\max_{Y} \min_{Z} f(Y,Z), \quad \begin{pmatrix} Y \\ Z \end{pmatrix} \in D$$

і мінімаксу

$$\min_{Y} \max_{Z} f(Y,Z), \quad \begin{pmatrix} Y \\ Z \end{pmatrix} \in D.$$

Зробимо зауваження щодо математичних методів дослідження операцій. Як математичні методи дослідження операцій зазвичай розуміють математичний апарат (спеціально розроблений або адаптований), що призначений для вирішення завдань дослідження операцій. Розробленість математичних методів для різних класів завдань дослідження операцій є далеко не однаковою. В даний час найбільш розробленою є теорія лінійного та опуклого програмування. Вкажемо основні напрями, за якими велися і ведуться математичні дослідження.

Розвиток симплекс-методу. В даний час існує багато модифікацій цього методу, що дозволяють істотно скоротити час розрахунку, зробити алгоритм нечутливим до виродженості опорних планів, підвисити розмірність вирішуваних задач, вирішувати так звані блокові задачі та ін. Не дивлячись на велику кількість цих модифікацій, продовжують з'являтися все нові й нові його варіанти.

Цілочислове лінійне програмування. У великій кількості вирішуваних задач лінійного програмування на змінні накладається додаткова умова їх цілочисловості. Коли накладено додаткову умову цілочисловості змінних, відповідне завдання носить назву завдання цілочислового лінійного програмування. Просте округлення до цілих чисел тут не допомагає, план може вийти не оптимальним. Тому доводиться розробляти спеціальні алгоритми рішення таких задач, до найвідоміших з яких належать алгоритми Гоморі, засновані на так званій ідеї відсікання.

Булеве програмування. До окремого випадку завдання цілочислового лінійного програмування належать завдання, де змінні можуть приймати лише два значення 0 та 1. Відповідні завдання часто називають завданнями булевого програмування. Найвідоміші з цих завдань – це завдання про призначення (якого працівника на яку роботу поставити), завдання вибору маршруту (завдання комівояжера, завдання листоноші), завдання про максимальне паросполучення та ін. Для вирішення завдань цього типу розробляються дуже специфічні алгоритми, засновані на комбінаториці, графах та ін.

Стохастичне лінійне програмування. Буває багато практичних ситуацій, коли коефіцієнти цільової функції, коефіцієнти в матриці коефіцієнтів, коефіцієнти обмежень є випадковими величинами. В цьому випадку сама цільова функція стає випадковою величиною, і обмеження типу нерівностей можуть виконуватися лише з певною імовірністю. Доводиться змінювати постановку самих завдань з урахуванням цих ефектів і розробляти абсолютно нові методи їх рішення.

Квадратичне програмування. Завдання лінійного програмування є окремим випадком цих завдань, вони виходять при A=0. Способи рішення цих задач багато в чому визначаються видом матриці A: якщо A - позитивно визначена матриця, то цільова функція буде опуклою і будь-який її локальний мінімум буде глобальним. Якщо A - негативно визначена матриця, то може бути кілька локальних мінімумів, але глобальний мінімум, якщо він існує, досягається обов'язково на вершині припустимої області. У загальному випадку, коли власні числа матриці A мають різні знаки, завдання дуже сильно ускладнюється, оскільки глобальний мінімум може досягатися де завгодно - і усередині області і на її межі.

Опукле програмування. Для цих завдань характерним є те, що будь-який локальний мінімум виявляється глобальним, і все збігається до знаходження цього єдиного мінімуму. Для вирішення завдань цього типу розроблені численні числові методи, пристосовані для вирішення на ЕОМ, в основному пов'язані з поняттям градієнту цільової функції і основною ідеєю про те, що функція найшвидше убуває, якщо рухатися в напрямі, протилежному градієнту. До них належать метод градієнтного спуску, метод пов'язаних градієнтів та ін. Але є і методи, засновані на інших ідеях - метод штрафних функцій, численні варіанти методу випадкового пошуку та ін.

Геометричне програмування. Під завданнями геометричного програмування розуміють завдання найбільш щільного розташування деяких об'єктів в заданій двовимірній або тривимірній області. Такі завдання зустрічаються в завданнях розкрою матеріалу для виробництва якихось виробів та ін. Це ще недостатньо розроблена область математичного програмування і алгоритми, що є тут, в основному орієнтовані на скорочення перебору варіантів з пошуком локальних мінімумів.

Дискретне програмування. Багато завдань дослідження операцій такі як розподіл ресурсів, сіткове планування, календарне планування описуються математичними моделями дискретного програмування. У завданнях дискретного програмування область припустимих рішень є неопуклою і незв'язною, тому відшукання рішення в таких завданнях пов'язане з істотними труднощами. Зокрема неможливе застосування стандартних прийомів, використовуваних при заміні дискретного завдання його безперервним аналогом, що полягають в подальшому округленні знайденого рішення до найближчого цілочислового.

Динамічне програмування - це обчислювальний метод для вирішення завдань певної структури. Виникло і сформувалося в 1950-1953 рр. завдяки роботам Р. Беллмана над динамічними завданнями управління запасами. У спрощеному формулюванні динамічне програмування є спрямований послідовний перебір варіантів, який обов'язково призводить до глобального максимуму. Основні необхідні властивості завдань, до яких можливо застосувати цей принцип, наступні:

- завдання має припускати інтерпретацію як n-кроковий процес ухвалення рішень;
- завдання має бути визначеним для будь-якої кількості кроків і мати структуру, не залежну від їх кількості;
- під час розгляду k-крокового завдання має бути заданою певна множина параметрів, що описують стан системи, від яких залежать оптимальні значення змінних. Причому ця множина не повинна змінюватися при збільшенні кількості кроків;
- вибір рішення (управління) на k-му кроці не повинен робити вплив на попередні рішення, окрім необхідного перерахунку змінних.

Завдання про вибір траєкторії, завдання послідовного ухвалення рішення, завдання про використання робочої сили, завдання управління запасами є класичними завданнями динамічного програмування.

Контрольні запитання

1. Що розуміють під дослідженням операцій?
2. Сформулюйте основне завдання дослідження операцій.
3. Чи існує зв'язок між критерієм оптимальності та принципом оптимальності? Відповідь аргументуйте.
4. У чому полягає принципова відмінність статичних і динамічних завдань дослідження операцій?
5. Наведіть класифікацію завдань дослідження операцій за структурою інформаційного стану ОПР.
6. Яке завдання дослідження операцій називають параметричним, в чому його принципова відмінність від завдання ухвалення рішень в умовах невизначеності?
7. Наведіть класифікацію завдань дослідження операцій за видом критерію оптимальності.
8. Який зв'язок існує між завданнями математичного програмування і векторної оптимізації?

ТЕМА 2
ОПТИМІЗАЦІЙНІ ЗАДАЧІ УПРАВЛІННЯ ЗАПАСАМИ

2.1 Основні поняття та визначення теорії управління запасами

Завдання управління запасами є одним з найчисленніших класів економічних завдань дослідження операцій, розв’язання яких має важливе народногосподарське значення. Правильне та своєчасне визначення оптимальної стратегії управління запасами та їхнім нормативним рівнем забезпечує вивільнення істотних оборотних коштів, що заморожені у запасах, і підвищує ефективність використання ресурсів підприємства.

Для забезпечення безперервності виробничого процесу в бізнесі та у виробництві актуальною є підтримка розумного обсягу запасів матеріальних ресурсів. Запаси зазвичай розглядаються як неминучі витрати, причому занадто низький рівень запасу призводить до дорогих зупинок виробництва, а занадто високий - до заморожування капіталу.

Розглянемо основні поняття, використовувані в моделях управління запасами.

Попит на продукт, що запасається, являє собою обсяг запасу, використовуваний в одиницю часу. Попит на підприємстві може бути детермінованим (у найпростішому випадку - постійним у часі) або випадковим. Випадковість попиту зумовлюється або випадковим моментом попиту, або випадковим обсягом попиту в детерміновані або випадкові моменти часу.

Поповнення складу може здійснюватися або періодично через певні інтервали часу, або в міру зниження рівня запасу до певного значення.

Обсяг замовлення. При періодичному поповненні та випадковому вичерпанні запасів обсяг замовлення може залежати від того стану, що має місце в момент розміщення замовлення. Замовлення зазвичай подається на ту саму величину при зниженні рівня запасу до заданого обсягу, називаного **точкою замовлення**.

Час доставки. В ідеалізованих моделях управління запасами припускається, що замовлене поповнення надходить на склад миттєво. В інших моделях передбачається затримка поставок на фіксований або випадковий інтервал часу.

Вартість поставки. Уважають, що вартість кожної поставки складається з двох компонентів – **витрат на оформлення замовлення** й **витрат на придбання запасу**. Витрати на оформлення замовлення є постійними витратами й не залежать від обсягу замовлення. Витрати на придбання визначаються вартістю одиниці ресурсу, що надходить, і залежать, найчастіше лінійно, від обсягу замовлення. Вартість одиниці ресурсу може бути постійною або може мати знижку, що залежить від обсягу замовлення.

Витрати на зберігання запасу – це витрати на утримання запасу на складі, вони включають відсоток на інвестований капітал, а також вартість зберігання запасу.

Штраф за дефіцит – це витрати, зумовлені відсутністю запасу необхідного ресурсу. Вони включають потенційні втрати прибутку. Відсутність запасу в потрібний момент призводить до збитків, що пов'язані з простоєм обладнання, неритмічністю виробництва, втратою довіри клієнтів.

Номенклатура запасу. У найпростіших моделях припускається, що на складі зберігається запас однотипних виробів або однорідного продукту. У складніших моделях розглядається багатономенклатурний запас.

Структура складської системи. Найповніше розроблені математичні моделі одиночного складу. Однак на практиці зустрічаються й складніші структури: ієрархічні системи складів з різними періодами поповнення й часом доставки замовлень, з можливістю обміну запасами між складами одного рівня ієрархії та ін.

Завдання управління запасами полягає у кількаразовому розміщенні й надходженні замовлень заданих обсягів певного продукту в певні моменти часу. У цьому змісті стратегія управління запасами полягає у визначенні оптимального розміру замовлення й моменту замовлення. Економічний розмір замовлення повинен мінімізувати сумарні витрати управління запасами, структура яких представлена на рисунку 2.1.

$$\left\{\begin{matrix}\text{Сумарні}\\ \text{витрати}\\ \text{управління}\end{matrix}\right\} = \left\{\begin{matrix}\text{Витрати на}\\ \text{придбання}\end{matrix}\right\} + \left\{\begin{matrix}\text{Витрати на}\\ \text{оформлення}\\ \text{заказу}\end{matrix}\right\} + \left\{\begin{matrix}\text{Витрати на}\\ \text{зберігання}\\ \text{заказу}\end{matrix}\right\} + \left\{\begin{matrix}\text{Втрати від}\\ \text{дефіциту}\\ \text{запасу}\end{matrix}\right\}$$

Рисунок 2.1 - Структура функції витрат управління запасами

Всі складові функції витрат управління виражають як залежності від обсягу замовлення та інтервалу часу між замовленнями. Інтервал часу між замовленнями залежить від типу системи управління запасами на підприємстві. Якщо передбачається періодичний контроль стану запасів, то момент надходження нового замовлення збігається з початком періоду. Якщо на підприємстві ведеться безперервний контроль стану запасу, то нові замовлення розміщаються, коли рівень запасу опускається до точки поновлення замовлення.

Таким чином, як критерій ефективності прийнятої стратегії управління запасами приймається функція витрат, що представляє собою сумарні витрати на зберігання та надходження продукту, що запасається, у тому числі втрати від псування продукту при зберіганні і його морального старіння, втрати прибутку від заморожування капіталу та ін., і витрати на штрафи.

Управління запасами полягає у відшуканні такої стратегії поповнення й витрати запасів за якої функція витрат приймає мінімальне значення. У моделях завдання управління запасами враховують стратегію, прийняту на підприємстві. Розрізняють детерміновані моделі - статичні й динамічні, і стохастичні моделі. У статичних моделях обсяг попиту є постійним у часі, у динамічних моделях обсяг попиту залежить від часу. У стохастичних моделях попит є випадковим.

Розглянемо статичні детерміновані моделі оптимізації запасів.

2.2 Класичне завдання економічного розміру замовлення

Статична детермінована модель без дефіциту. Класична модель управління запасами припускає, що попит постійний у часі, запас поповнюється миттєво, дефіцит запасу відсутній. Позначимо обсяг замовлення n в одиницях ресурсу, інтенсивність попиту r в одиницях ресурсу на одиницю часу, тривалість циклу t_0 в одиницях часу. Замовлення обсягом n одиниць розміщується та поповнюється миттєво в момент, коли запас ресурсу стає рівним нулю. Зміну запасу в часі в класичній моделі представлено на рисунку 2.2.

Оскільки інтенсивність попиту r постійна, можна визначити тривалість циклу t_0

$$t_0 = \frac{n}{r}. \qquad (2.1)$$

Оскільки зниження запасу в часі відбувається лінійно, що пов'язано з постійним попитом, середній рівень запасу становить $\overline{n} = n/2$.

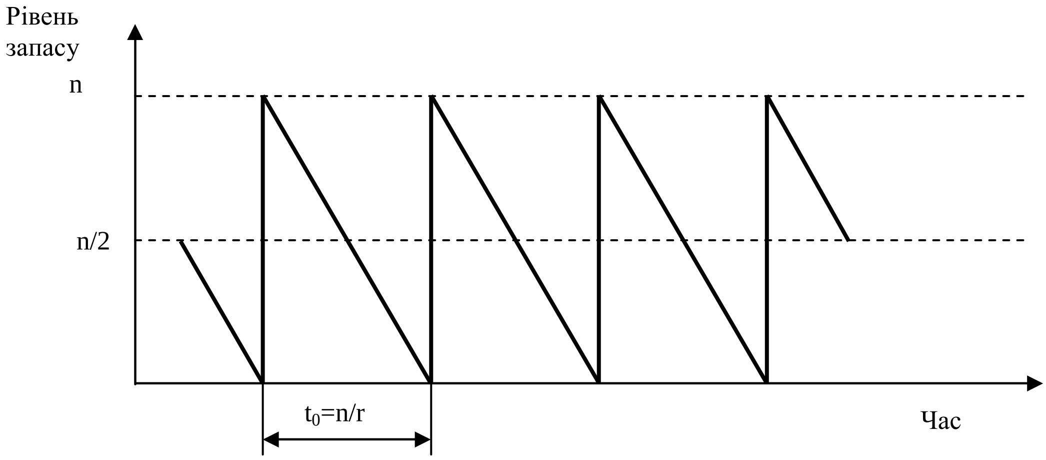


Рисунок 2.2 - Зміна запасу n у часі

Сформуємо функцію витрат управління. Нехай витрати на оформлення замовлення становлять c_1 грошових одиниць, а витрати на зберігання – c_2 грошових одиниць за одиницю продукції в одиницю часу. Тоді сумарні витрати в одиницю часу складуть

$$C(n) = \frac{c_1}{t_0} + c_2 * \frac{n}{2} = \frac{c_1}{n/r} + c_2 * \frac{n}{2}. \qquad (2.2)$$

Потрібно визначити оптимальний обсяг замовлення n*, тобто обсяг замовлення, що мінімізує витрати *C*(*n*). Візьмемо похідну від виразу сумарних витрат (2.2) і дорівняємо її до нуля:

$$\frac{dC(n)}{dn} = \frac{c_1 r}{n^2} + \frac{c_2}{2} = 0, \qquad (2.3)$$

звідки оптимальний обсяг замовлення *n**

$$n^* = \sqrt{\frac{2c_1 r}{c_2}}. \qquad (2.4)$$

Формула (2.4) називається формулою Уілсона, або формулою найбільш економічного обсягу запасу. Відповідно до (2.4) оптимальна стратегія формулюється в такий спосіб: замовляти $n^* = \sqrt{\dfrac{2c_1 r}{c_2}}$ одиниць продукції через кожні $t^*_0 = \dfrac{n^*}{r}$ одиниць часу.

Графіки складових функції витрат $C(n) = \dfrac{c_1 r}{n} + \dfrac{c_2 n}{2}$ наведені на рисунку 2.3. У точці мінімуму $C(n)$ її похідна дорівнює нулю

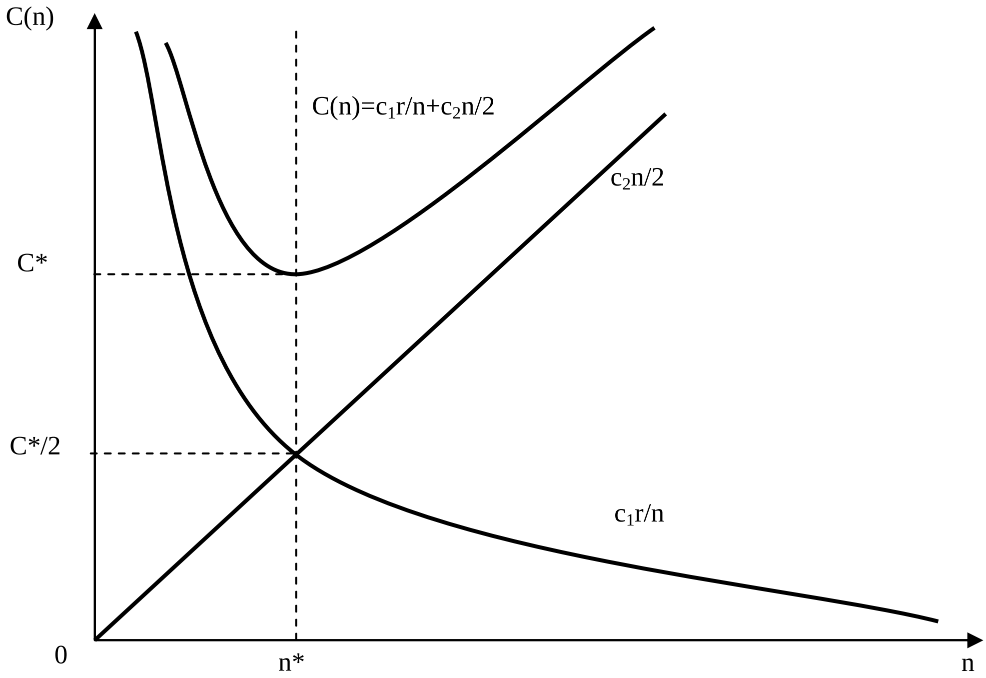


Рисунок 2.3 - Графіки функцій витрат

Статична детермінована модель із запізнюванням. У реальному завданні поповнення запасів не може відбутися миттєво. Між розміщенням замовлення та реальним його надходженням існує запізнювання $L<t^*_0$ (рис. 2.4).

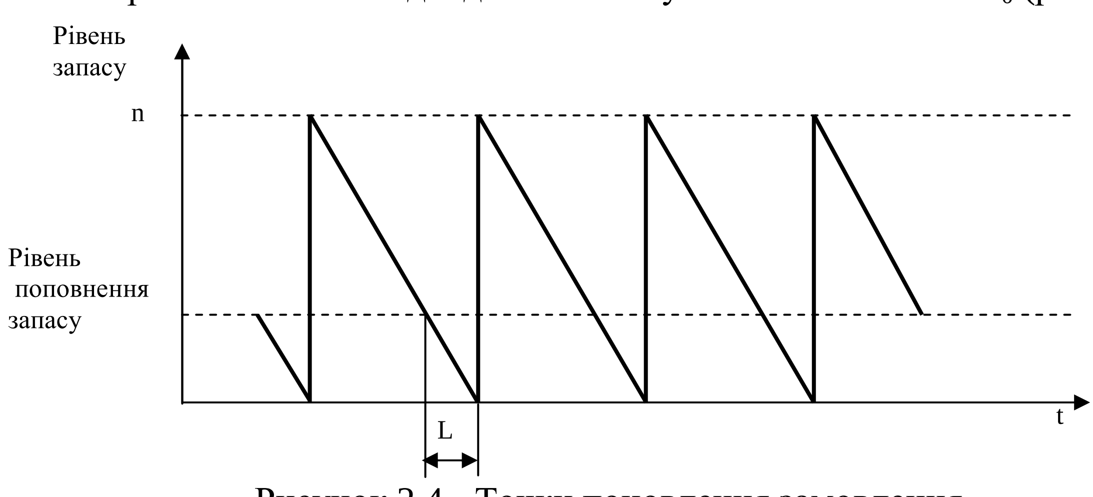


Рисунок 2.4 - Точки поновлення замовлення

З рисунку 2.4 видно, що точка поновлення замовлення має місце, коли рівень запасу знижується до $L_г$ одиниць. Якщо умова $L<t^*_0$ не виконується, то розраховують ефективний термін виконання замовлення $L_{еф}$. Для цього визначають найбільше ціле $k \le \frac{L}{t^*_0}$, і ефективний термін виконання замовлення $L_{еф}$ обчислюють за формулою

$$L_{эф} = L - k * t^*_0 .$$

Таким чином, точка поновлення замовлення має місце при рівні запасу

$$L_{еф}r \text{ одиниць}, \tag{2.5}$$

і оптимальна стратегія формулюється в такий спосіб: замовляти n^* одиниць продукції, коли рівень запасу знижується до $L_{еф}r$ одиниць.

Приклад 2.1. Попит r=100 одиниць/день, вартість замовлення становить c_1=600 грн., вартість зберігання однієї одиниці на складі c_2=0,02 грн./день, реальний термін виконання замовлення L= 12 днів. Визначити оптимальну стратегію управління запасами.

Розв'язання

Визначимо оптимальний обсяг замовлення за формулою (2.4)

$$n^* = \sqrt{\frac{2*100*100}{0,02}} = 1000 \text{ одиниць.}$$

Визначимо тривалість циклу за формулою (2.1)

$$t^*_0 = \frac{1000}{100} = 10 \text{ днів.}$$

Оскільки L=12 днів $>t^*_0$ =10 днів, розрахуємо ефективний термін виконання замовлення $L_{еф}$. Визначимо найбільше ціле $k \le \frac{12}{10}$, k=1, тоді ефективний термін виконання замовлення

$$L_{еф} = 12 - 1*10 = 2 \text{ дні} .$$

Отже, точка поновлення замовлення становить $L_{еф}r$=2*100=200 одиниць. Оптимальна стратегія формулюється в такий спосіб: замовити 1000 одиниць при зниженні запасу до 200 одиниць. Дана стратегія забезпечить мінімальні сумарні витрати на день, розмір яких відповідно до формули (2.2) складе

$$C(n) = \frac{100}{10} + 0,02*\frac{1000}{2} = 10 + 10 = 20 \text{ грн./ день}$$

Статична детермінована модель із розривами цін. Дана модель відрізняється тим, що продукт, що запасається, може бути придбаний із знижкою, що залежить від обсягу замовлення. Нехай вартість одиниці продукту без знижки становить $c_б$, а із знижкою - $c_з$. Причому, знижка надається, якщо обсяг замовлення n перевищує певний фіксований рівень q. Тоді вартість одиниці продукту c визначиться в такий спосіб:

$$c = \begin{cases} c_б, & \text{якщо} \quad n \le q, \\ c_з, & \text{якщо} \quad n > q, \end{cases}$$

де $c_б > c_з$.

Тоді витрати на придбання продукту в одиницю часу складуть

$$c=\begin{cases}\dfrac{c_б n}{t_0}=\dfrac{c_б n}{n/r}=c_б r, & якщо \quad n\le q,\\ \dfrac{c_з n}{t_0}=\dfrac{c_з n}{n/r}=c_з r, & якщо \quad n>q.\end{cases}$$

Функція сумарних витрат $C(n)$ прийме вигляд

$$C(n)=\begin{cases}C_б(n)=c_б r+\dfrac{c_1 r}{n}+c_2*\dfrac{n}{2}, & n\le q,\\ C_з(n)=c_з r+\dfrac{c_1 r}{n}+c_2*\dfrac{n}{2}, & n>q.\end{cases} \qquad (2.6)$$

Оскільки значення функцій $C_б(n)$ і $C_з(n)$ відрізняються на постійну величину $C_б(n)$-$C_з(n)$, точки їхнього мінімуму збігаються й відповідно до (2.4) знаходяться у точці $n_m=\sqrt{\dfrac{2c_1 r}{c_2}}$ (рис. 2.5). Мінімуму сумарних витрат без знижки на кривій $C_з(n)$ відповідає точка з абсцисою M.

З рисунку 2.5 видно, що оптимальний обсяг замовлення n^* залежить від того, де знаходиться точка розриву ціни q стосовно інтервалів $[0, n_m)$, $[n_m, M)$ і $[M,\infty)$. Величину M можна визначити з рівності

$$C_c(M)=C_б(M),$$

звідки

$$c_з r+\frac{c_1 r}{M}+c_2*\frac{M}{2}=c_б r+\frac{c_1 r}{n_m}+c_2*\frac{n_m}{2}. \qquad (2.7)$$

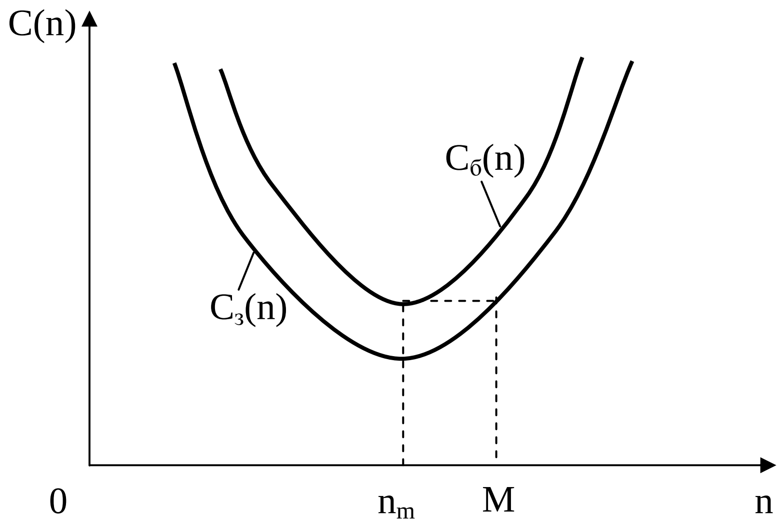


Рисунок 2.5 - Функції витрат без знижки та із знижкою

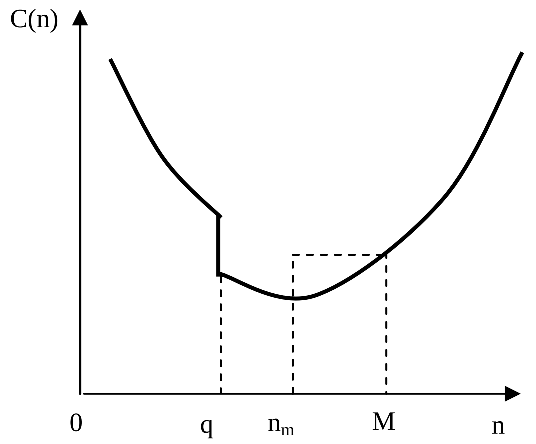
Вираз (2.7) після перетворень дає квадратне рівняння щодо M:

$$M^2+\frac{2c_з r-C_б(n_m)}{c_2}M+\frac{2c_1 r}{c_2}=0. \qquad (2.8)$$

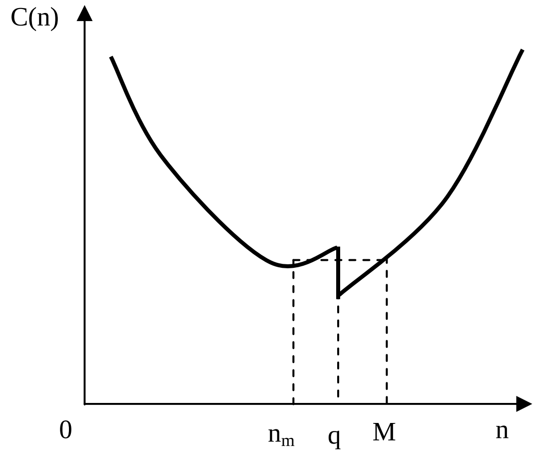
Припустимо, що за умовами поставки q знаходиться в інтервалі $[0, n_m)$.

У цьому випадку треба обчислити $n_m = \sqrt{\dfrac{2c_1 r}{c_2}}$ та прийняти $n^*=n_{\text{m}}$.

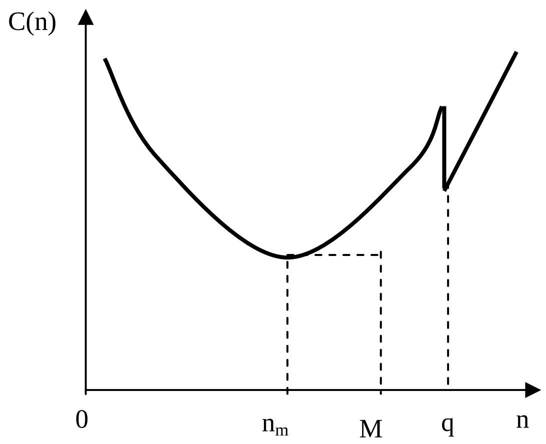
Якщо q не попадає в інтервал $[0,\ n_{\text{m}})$, потрібно обчислити M з рівняння (2.8) і визначити інтервали $[n_{\text{m}},M)$ і $[M,\infty)$. Якщо q знаходиться в інтервалі $[n_{\text{m}},M)$, приймаємо $n^*=q$. Якщо q знаходиться в інтервалі $[M,\infty)$, приймаємо $n^*=n_{\text{m}}$. Відповідні криві показані на рисунку 2.6.



а



б



в

Рисунок 2.6 - Функції витрат із знижкою

Очевидно, що якщо обсяг замовлення q, зумовлений знижкою, потрапить в інтервал $[0,\ n_{\text{m}})$, значення функції сумарних витрат, що лежать правіше q, зменшуються, і потрібно приймати $n^*=n_{\text{m}}$ (рис. 2.6, *а*). Якщо q попадає в інтервал $[n_{\text{m}},M)$, значення функції сумарних витрат, що лежать правіше q, збільшуються, і треба приймати $n^*=q$ (рис. 2.6, *б*). Якщо q знаходиться в інтервалі $[M,\infty)$, використовувати знижку не вигідно, і треба прийняти $n^*=n_{\text{m}}$ (рис. 2.6, *в*).

Приклад 2.2. АЗС займається заміною олії в автомобілях. За день обслуговується близько 150 автомобілів, кожний з яких вимагає доливки 1,25 л олії. Вартість зберігання олії становить 0,02 грн. за літр за один день. Покупна ціна олії - 3 грн. за літр. При покупці понад 1000 л надається знижка, і тоді 1 л олії коштує 2,50 грн. Вартість розміщення замовлення становить 20 грн., термін виконання замовлення 2 дні. Визначити оптимальну стратегію управління запасами.

Розв’язання

Одноденна потреба в олії становить 150*1,25=187,5 л.; c_1=20 грн.; c_2=0,02 грн. за літр у день; L= 2 дні; $c_б$=3 грн. за літр; $c_з$=2,50 грн. за літр; q=1000 літрів.

Обчислимо мінімум функції сумарних витрат, для чого визначимо n_m

$$n_m = \sqrt{\frac{2c_1 r}{c_2}} = \sqrt{\frac{2*20*187,5}{0,02}} = 612,37\ л.$$

Оскільки q=1000 > n_m=612,37, обчислимо точку M

$$C_б(n) = c_б r + \frac{c_1 r}{n_m} + c_2 * \frac{n_m}{2} = 3*187,5 + 612,37 + 0,02*\frac{612,37}{2} = 574,75,$$

$$M^2 + \frac{2*2,5*187,5 - 574,5}{0,02} M + \frac{2*20*187,5}{0,02} = 0,$$

$$M^2 - 10599,74M + 375000 = 0,$$

звідки дістанемо M=10564,5. Другий інтервал [612,37, 10564,5), q=1000 лежить у другому інтервалі, отже $n^*=q$= 1000 л.

Точка поновлення замовлення при L= 2 дні становить 2*187,5=375 л.

Оптимальна стратегія формулюється в такий спосіб. Замовити 1000 л олії, коли рівень запасу зменшиться до 375 літрів.

2.3 Статична детермінована модель з дефіцитом

Дефіцит означає, що за відсутності запасу інтенсивність попиту r залишається тією самою, і підприємство через відсутність продукту зазнає збитків c_3. Графік моделі з дефіцитом наведений на рисунку 2.7. Необхідний рівень запасу для виробництва становить n одиниць. Він витрачається з попитом r одиниць продукту в одиницю часу. Однак оскільки обсяг запасу продукту становить s одиниць, виникає дефіцит n-s.

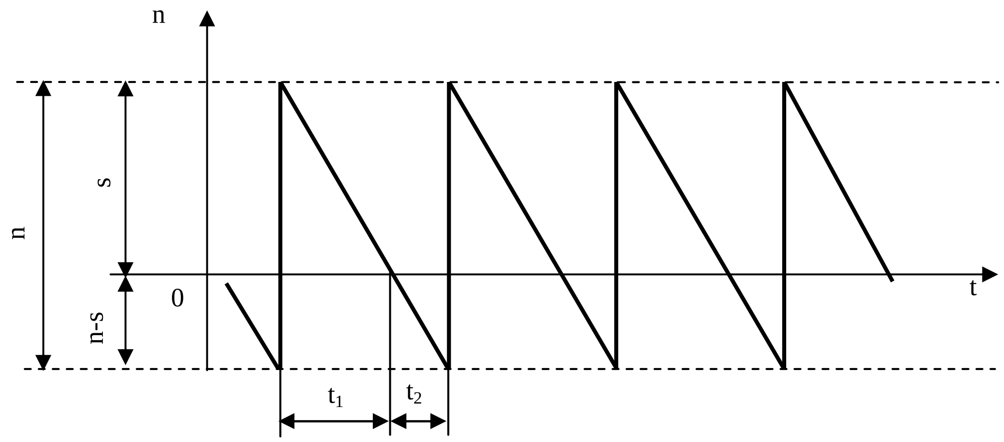


Рисунок 2.7 - Нагромадження дефіциту

Запас продукту s витрачається з попитом r за час

$$t_1 = \frac{s}{r}. \tag{2.9}$$

За час

$$t_2 = \frac{n-s}{r} \tag{2.10}$$

накопичується дефіцит *n-s*.

Складові сумарної функції витрат *C*(*n,s*) визначимо в такий спосіб. Витрати на замовлення запасу в одиницю часу, як і в попередньому випадку, складуть $\frac{c_1}{t_0} = \frac{c_1 r}{n}$. Витрати на зберігання середнього числа одиниць продукту $\bar{s} = \frac{s}{2}$ при лінійній витраті складуть в одиницю часу

$$c_2 * \frac{s}{2} * t_1 = c_2 * \frac{s}{2} * \frac{s}{r} = c_2 * \frac{s}{2} * \frac{s}{n} t_0 = c_2 * \frac{s^2}{2n} * t_0 .$$

Штраф за час t_2 складе $c_3 * \frac{n-s}{2} * \frac{n-s}{r} = c_3 * \frac{(n-s)^2}{2n} t_0$. Тоді сумарна функція витрат в одиницю часу прийме вигляд

$$C(n,s) = \frac{c_1 r}{n} + c_2 * \frac{s^2}{2n} * t_0 + c_3 * \frac{(n-s)^2}{2n} t_0 . \tag{2.11}$$

Дане завдання вимагає відшукання такого обсягу запасу *n** і максимального рівня запасу *s**, за яких функція витрат мінімальна.

Покажемо, що при *n*=*s* вираз (2.11) збігається з (2.2):

$$C(n) = \frac{c_1 r}{n} + c_2 * \frac{n^2}{2n} * t_0 + c_3 * \frac{(n-n)^2}{2n} t_0 = \frac{c_1 r}{n} + c_2 * \frac{n}{2} * t_0 .$$

Дослідимо функцію *C*(*n*,*s*) на екстремум, для чого візьмемо часткові похідні та дорівняємо їх нулю, одержимо

$$\begin{cases} \frac{\partial C(n,s)}{\partial n} = -\frac{c_1 r}{n^2} - c_2 * \frac{s^2}{2n^2} * t_0 + c_3 * \frac{1}{2} t_0 - c_3 * \frac{s^2}{2n^2} t_0 = 0, \\ \frac{\partial C(n,s)}{\partial s} = c_2 * \frac{s}{n} * t_0 - c_3 * t_0 + c_3 * \frac{s}{n} t_0 = 0. \end{cases}$$

Зробимо перетворення. Помножимо перше рівняння на $2n^2$ і розділимо на t_0, друге рівняння помножимо на n і розділимо на t_0, одержимо

$$\begin{cases} c_3 n^2 - s^2 (c_2 + c_3) = \frac{2c_1 r}{t_0}, \\ s = n \frac{c_3}{c_2 + c_3}. \end{cases} \tag{2.12}$$

З (2.12) виразимо $\tilde{n}*$, підставимо для цього вираз *s* з другого рівняння в перше:

$$c_3 n^2 - n^2 \frac{c_3^2}{(c_2 + c_3)^2} (c_2 + c_3) = \frac{2c_1 r}{t_0} ,$$

$$c_3 n^2 \frac{c_2}{c_2 + c_3} = \frac{2c_1 r}{t_0} ,$$

$$n^2 \frac{c_2 c_3}{c_2 + c_3} = \frac{2c_1 r}{t_0} .$$

Остаточно одержимо

$$\tilde{n}^* = \sqrt{\frac{2c_1 r}{c_2 t_0}} \sqrt{\frac{c_2 + c_3}{c_3}}, \qquad (2.13)$$

$$\tilde{s}^* = \sqrt{\frac{2c_1 r}{c_2 t_0}} \sqrt{\frac{c_3}{c_2 + c_3}} = \tilde{n}^* \frac{c_3}{c_2 + c_3}, \qquad (2.14)$$

де коефіцієнт $\rho = \dfrac{c_3}{c_2 + c_3}$ являє собою щільність збитків через незадоволений попит. Даний показник приймає значення в інтервалі $0 \le \rho \le 1$.

Якщо значення штрафу за дефіцит c_3 порівняно з витратами на зберігання продукту c_2 малим, то ρ близький до нуля. Якщо штраф за дефіцит c_3 істотно перевищує витрати на зберігання продукту c_2, то ρ наближається до одиниці. Використовуючи коефіцієнт ρ, формули (2.13) і (2.14) можна записати в такий спосіб

$$\tilde{n}^* = \sqrt{\frac{2c_1 r}{c_2 \rho}}, \qquad (2.15)$$

$$\tilde{s}^* = \tilde{n}\rho. \qquad (2.16)$$

З формули (2.16), використовуючи формули (2.9) і (2.10), можна одержати співвідношення

$$\frac{t_1}{t_0} = \frac{\tilde{s}^*}{\tilde{n}^*} = \rho \quad \text{і} \quad \frac{t_2}{t_0} = \frac{\tilde{n}^* - \tilde{s}^*}{\tilde{n}^*} = 1 - \rho. \qquad (2.17)$$

Таким чином, коефіцієнт ρ показує частку часу, протягом якого запас продукту буде відсутній. Вона становить $(1-\rho)t_0$.

Оптимальні обсяги замовлення для моделей з дефіцитом і без дефіциту зв'язані співвідношенням

$$\tilde{n}^* = \frac{n^*}{\sqrt{\rho}}. \qquad (2.18)$$

З виразу (2.18) випливає, що оптимальний обсяг замовлення в завданні з дефіцитом перевищує оптимальний обсяг замовлення в завданні без дефіциту в $\dfrac{1}{\sqrt{\rho}}$ разів.

2.4 Стохастичні моделі управління запасами

Стохастичними моделями управління запасами є такі моделі, у яких попит є випадковою величиною, що істотно ускладнює їхній аналіз.

Припустимо, що попит r на інтервалі часу Т є випадковим. Відомий закон розподілу попиту, що заданий у вигляді ряду розподілу $p(r)$, функції розподілу $F(r)$ або щільності розподілу ймовірностей $f(r)$. Якщо попит r не перевищує рівня запасу s, то придбання та зберігання надлишку продукту вимагає додаткових витрат c_2 на одиницю продукту. Якщо попит r перевищує рівень запасу s, то це призводить до штрафу за дефіцит c_3 на одиницю продукту.

Функція сумарних витрат у стохастичних моделях є випадковою величиною, тому використовують її імовірнісні характеристики. Якщо випадковий попит r є дискретною випадковою величиною і його закон розподілу $p(r)$, то математичне сподівання функції сумарних витрат $C(s)$ визначається виразом:

$$C(s) = c_2 \sum_{r=0}^{s} (s-r)\,p(r) + c_3 \sum_{r=s+1}^{\infty} (r-s)\,p(r) \quad . \qquad (2.19)$$

У виразі (2.19) перший доданок ураховує витрати на придбання та зберігання надлишку s-r одиниць продукту при $r < s$, а другий доданок - штраф за дефіцит, що дорівнює r-s одиниць продукту при $r > s$.

Якщо випадковий попит є безперервною випадковою величиною, і його закон розподілу заданий щільністю розподілу ймовірностей $f(r)$, то математичне сподівання функції сумарних витрат $C(s)$ визначається виразом:

$$C(s) = c_2 \int_0^s (s-r)\,f(r)dr + c_3 \int_0^s (r-s)\,f(r)dr \quad . \qquad (2.20)$$

Завдання оптимізації управління запасами полягає у відшуканні такого запасу s*, за якого математичне сподівання сумарних витрат $C(s)$ приймає мінімальне значення.

Можна показати, що при дискретному випадковому попиті r функція сумарних витрат приймає мінімальне значення при такому s*, при якому функція розподілу попиту $F(r)=P\{r<s\}$ задовольняє нерівностям

$$F(s^*) < \rho < F(s^*+1). \qquad (2.21)$$

При безперервному випадковому попиті r значення сумарних витрат $C(s)$ приймає мінімум при значенні s*, яке визначається з рівняння

$$F(s^*)= \rho, \qquad (2.22)$$

де ρ - щільність збитків через незадоволений попит, що визначається за формулою

$$p = \frac{c_3}{c_2 + c_3}.$$

Оптимальний запас s* при безперервному попиті, що відповідає знайденому значенню ρ, можна визначити графічно (рис. 2.8).

Приклад 2.3. Підприємство в міру надходження заявок провадить закупівлю агрегату із запасними блоками до нього. Вартість одного блоку дорівнює 5 грн. У випадку виходу агрегату з ладу через поломку блоку, відсутнього в запасі, простій агрегату і термінове замовлення нового блоку до нього обійдеться в 100 грн. Статистичний розподіл числа блоків, що потребують заміну, подано в таблиці 2.1.

Визначимо оптимальне число запасних блоків, яке потрібно придбати разом з агрегатом.

Таблиця 2.1 - Ряд розподілу числа блоків, що потребують заміну

Число замінених блоків r	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Статистична ймовірність агрегатів $p(r)$, яким потрібна була заміна r блоків	0,90	0,05	0,02	0,01	0,01	0,01	0,00

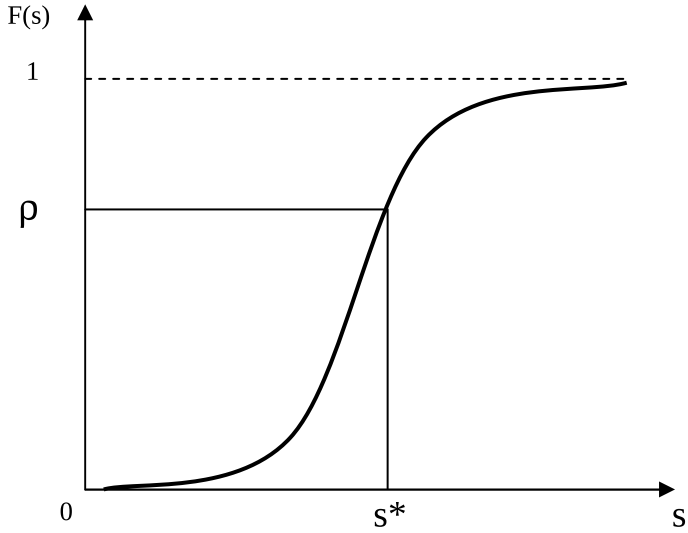


Рисунок 2.8 - Графічний пошук оптимального запасу

Розв’язання

Відомі значення c_2=5 грн. і c_3=100 грн. Обчислимо щільність збитків через недостачу запасних блоків

$$\rho = \frac{c_3}{c_2 + c_3} = \frac{100}{5+100} = 0{,}952$$

Визначимо значення функції розподілу попиту $F(r)=P(r<s)$ і зведемо їх у таблицю 2.2.

Таблиця 2.2 - Функція розподілу попиту r

s	0	1	2	3	4	5	6	>6
r	0	1	2	3	4	5	6	>6
***F*(*s*)**	0	0,9	0,95	0,96	0,97	0,98	0,99	1,00

З таблиці 2.2 видно, що оптимальний запас становить $s^*=2$, оскільки він задовольняє нерівність

$$F(2) < 0{,}952 < F(3).$$

Отже, під час покупки агрегату необхідно покупати до нього два запасних блоки.

Контрольні запитання

1. Поясніть, що у задачах управління запасами розуміють як попит?
2. Поясніть складові вартості поставки продукту, що запасається.
3. Які витрати включають до штрафу за дефіцит?
4. На яких припущеннях ґрунтується класичне завдання економічного розміру замовлення?
5. Запишіть формулу Уілсона та поясніть її складові.
6. Поясніть графіки функцій витрат на рисунку 2.3.
7. Як відрізняються статичні та динамічні, детерміновані та стохастичні моделі?
8. Як у завданні економічного розміру замовлення враховується знижка

на продукт, що запасається?

9. На яких припущеннях ґрунтується статична детермінована модель з дефіцитом?

10. Запишіть формулу коефіцієнту ρ та поясніть його економічний зміст.

11. У якому випадку використовують стохастичні моделі управління запасами?

12. Охарактеризуйте імовірнісні характеристики випадкового попиту та оптимального запасу у стохастичній моделі управління запасами.

ТЕМА 3
ЗАДАЧІ УПОРЯДКУВАННЯ ТА КООРДИНАЦІЇ

3.1 Функції координації та регулювання як функції управління

Координація є функцією впорядкування діяльності. Вона спрямована на забезпечення узгодженості дій для успішного досягнення цілі. Дана функція може реалізовуватися в різних формах: уточнення функцій; узгодження цілей і завдань, планів і заходів. Координація є одним з основних засобів підвищення ефективності діяльності в будь-якій сфері, оскільки дозволяє сконцентрувати зусилля на необхідному напрямі, уникнути паралелізму і дублювання. Не дивлячись на широкий діапазон поглядів на місце координації в управлінні, з урахуванням всіх класифікаційних ознак, є підстави для виділення координації як самостійної функції управління, оскільки вона має свою специфічну ціль - підтримка пропорційності в діяльності системи. Вона проявляється на всіх стадіях процесу управління.

В ході будь-якої діяльності неминуче виникають відхилення, які викликаються, з одного боку, змінами ситуацій, з іншого - недоліками в діяльності як об'єкту, так і суб'єкту управління. В зв'язку з цим необхідна відповідна функція для амортизації негативних явищ. Цю роль виконує функція регулювання. Функція регулювання спрямована на підтримку динамічної рівноваги системи в ході забезпечення її життєдіяльності. Вона сприяє адаптації системи до ситуацій, що постійно змінюються, попередження і усунення відхилень від запланованих заходів.

Функцію регулювання необхідно відрізняти від функції організації, до якої її іноді включають. Останнє пояснюється схожістю цілей цих функцій. Основна відмінність полягає в тому, що організація забезпечує створення і підготовку системи для виконання поставлених завдань, встановлення необхідних для цього відносин між структурними одиницями, а регулювання спрямоване на своєчасне коректування дій виконавців, суб'єкта управління. В результаті регулювання зовнішні і внутрішні чинники, що дезорганізують діяльність, нейтралізуються. Велику роль в регулюванні грає зв'язок між суб'єктом і об'єктом управління. Добре налагоджена система інформації дозволяє не тільки усувати відхилення в діяльності, але і попереджати їх. Регулювання здійснюється в основному шляхом розпорядливої діяльності суб'єкта у формі наказів, вказівок, розпоряджень.

Отже, регулювання як функція управління - вид діяльності менеджера по усуненню виявлених відхилень від нормального режиму роботи організації. Таким чином, координація і регулювання як функції управління істотного доповнюють функцію організації, забезпечуючи взаємодію окремих частин системи управління на користь виконання завдань, що стоять перед нею.

3.2 Характеристика задач упорядкування та координації

Задачі упорядкування та координації виникають при необхідності визначення оптимального управління окремими операціями або комплексами операцій. На відміну від завдань масового обслуговування в цих завданнях визначаються оптимальні послідовність, порядок виконання робіт, терміни початку і закінчення кожної операції, об'єми ресурсів, що виділяються. Оптимальними вважаються рішення, що забезпечують найбільшу ефективність системи в цілому, тобто оптимальне значення прийнятого критерію.

Багато завдань з організації виробництва, календарного планування в їх математичному трактуванні належать до задач упорядкування. Класичним прикладом задачі упорядкування, що знаходить застосування в різних галузях промисловості, є завдання з планування виробничої лінії. Ця задача упорядкування формулюється так: визначити порядок запуску деталей у виробництво, що мінімізує сумарний час виконання плану.

Характерними задачами упорядкування є задачі узгодження (сіткового планування), масового обслуговування, вибору маршрутів, змагальних (завдання теорії ігор), управління запасами та ін. Для більшості задач упорядкування рішення доводиться шукати за допомогою методу моделювання або через евристичні алгоритми побудови рішення.

Одним з методів оптимізації задач упорядкування перебору варіантів є метод динамічного програмування. Цей метод на відміну від методів лінійного програмування не залежить від характеру цільової функції і не вимагає лінійності початкових залежностей. Інші методи – це методи комбінаторної оптимізації, яка полягає в пошуку оптимального об'єкту в кінцевій множині об'єктів, чим комбінаторна оптимізація дуже схожа на дискретне програмування. Під дискретним програмуванням розуміють цілочислове програмування, а комбінаторна оптимізація має справу з графами та іншими структурами. Проте дискретне програмування та комбінаторна оптимізація дуже близько зв'язані і в літературі часто переплітаються. Комбінаторна оптимізація часто збігається до визначення ефективного розподілу ресурсів, використовуваних для пошуку оптимального рішення.

Комбінаторна оптимізація включає завдання оптимізації, в яких множина припустимих рішень є дискретною, або її можна звести до дискретної множини. У багатьох завданнях комбінаторної оптимізації повний перебір варіантів нереальний. Але існують алгоритми, що працюють на деяких класах завдань дискретного програмування, істотна частина яких належить до теорії лінійного програмування. Приклади комбінаторної оптимізації, що потрапляють в цю область - це завдання пошуку найкоротшого шляху і дерева найкоротших шляхів, визначення максимального потоку, знаходження остовних дерев, знаходження паросполучень.

Завдання комбінаторної оптимізації можна розглядати як пошук кращого елементу в деякій дискретній множині, тому, у принципі, можуть бути використані будь-які алгоритми пошуку або метаеврістичні алгоритми. Проте загальні алгоритми пошуку не гарантують ні оптимального рішення, ні швидкого рішення. Загальним алгоритмічним методом для знаходження оптимальних рішень різних задач оптимізації, особливо дискретної і комбінаторної оптимізації, є метод віток і границь. Цей метод є комбінаторним, він являє собою алгоритм перебору з відсівом підмножин множини припустимих рішень, що не містять оптимальних рішень. Метод був вперше запропонований Ленд і Дойг в 1960 р. для вирішення завдань цілочислового лінійного програмування.

Частковими завданнями комбінаторної оптимізації є наступні:

- завдання вибору маршруту (завдання комівояжера);
- завдання про ранець;
- завдання розкрою;
- завдання про призначення.

Завдання комівояжера є одним з найвідоміших завдань комбінаторної оптимізації. Завдання полягає у відшуканні найвигіднішого маршруту, що проходить через вказані міста хоч би по одному разу з подальшим поверненням в початкове місто. В умовах завдання вказуються критерій ефективності маршруту (найкоротший, найдешевший, сукупний критерій та ін.) і відповідні матриці відстаней, вартості та ін. Як правило, маршрут повинен проходити через кожне місто тільки один раз.

Завдання про ранець назву одержало від завдання максимізації укладання як можна більшого числа потрібних речей в рюкзак за умови, що загальний об'єм (або вага) всіх предметів обмежений. Подібні завдання часто виникають в економіці, прикладній математиці, криптографії. У загальному вигляді, завдання можна сформулювати так: з необмеженої множини предметів з властивостями «вартість» і «вага», потрібно відібрати певне число предметів так, щоб одержати максимальну сумарну вартість при одночасному дотриманні обмеження на сумарну вагу.

У загальному вигляді задачу цілочислового програмування можна сформулювати як задачу знаходження максимуму (або мінімуму) цільової функції $L(x_1, x_2,..., x_n)$ на множині D, що зумовлена системою обмежень. Загальну задачу цілочислового програмування формулюють в такий спосіб:

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \rightarrow extr \qquad (3.1)$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j = b_i; \quad i = \overline{1,m} \tag{3.2}$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,n}.$$

Умову $x_j - цілі$ називають умовою дискретності.

Однією з найбільш простих в смислі постановки і найбільш важких щодо отримання рішення в загальному вигляді задач упорядкування є завдання Джонсона. Сформулюємо завдання Джонсона з двома верстатами.

Є n деталей і два верстати. Кожна деталь повинна спочатку пройти обробку на першому верстаті, потім - на другому. При цьому i-а деталь обробляється на першому верстаті за a_i часу, а на другому - за b_i часу. Кожен верстат в кожен момент часу може працювати тільки з однією деталлю. Потрібно скласти такий порядок подачі деталей на верстати, щоб підсумковий час обробки всіх деталей був би мінімальним.

Це завдання іноді називається завданням двопроцесорного обслуговування завдань, або завданням Джонсона (на ім'я S.M. Johnson, який в 1954 р. запропонував алгоритм для її вирішення).

Окремими випадками комбінаторних завдань є сукупність моделей календарного планування і розроблених для їх вирішення методів сіткового планування та управління. Зазвичай завдання ставиться таким чином: скласти план виготовлення всіх виробів, в якому не порушувалися б технологічні обмеження, обмеження з потужності обладнання, а також терміни запуску і випуску продукції.

3.3 Сіткова модель. Основні поняття та визначення

В рамках теорії дослідження операцій розглядається множина практичних завдань, які можна сформулювати як сіткові моделі, наприклад, складання часового графіка будівельних робіт. На основі сіткових моделей розроблено низку методів планування, складання розкладів і управління проектами.

Сіткове планування та управління (СПУ) являє собою комплекс спеціальних графічних і розрахункових методів, що призначені для моделювання і аналізу планів реалізації складних проектів і розробок, наприклад, таких як розробка туристської послуги, або дослідження системи управління деякою організацією, або маркетингове дослідження, та ін. Найбільше розповсюдження дістали метод критичного шляху (Critical Path Method або CPM) і система планування і управління програмами розробок (Program Evaluation and Review Technique або PERT). Ці методи розглядають проект як сукупність взаємозв'язаних процесів, що вимагають часових та інших ресурсів. Як процес розуміють вид діяльності, етап або фазу виконання проекту. Методи CPM і PERT відрізняються тим, що в методі критичного шляху CPM тривалість кожного етапу проекту є детермінованою, а в методі PERT - стохастичною.

СПУ призначено для розроблення управлінських рішень, що спрямовані на дотримання хронологічного графіка робіт, або на зміну графіка робіт відповідно до реальних можливостей, якщо в ході виконання проекту з'ясовується, що деякі роботи неможливо виконати в запланований термін.

Під проектом розуміють комплекс заходів, у результаті виконання яких на певний термін при обмежених матеріальних ресурсах має бути досягнута заздалегідь визначена система цілей. Характерною рисою таких проектів є те, що вони складаються з низки окремих робіт. Роботи пов'язані одна з одною так, що виконання деяких з них не можна розпочати раніше, ніж будуть завершені деякі інші роботи. Наприклад, розрахувати ціну послуги неможливо раніше, ніж буде складена калькуляція; реалізація нового тура не може бути здійснена, якщо ще не навчений персонал, та ін.

З позицій наукового менеджменту процес планування та управління проектами можна розглядати з різних рівнів. З погляду стратегічного управління фірмою, у рамках якої реалізується проект (горизонт планування при цьому становить кілька років); з погляду оперативного управління, що орієнтований на часовий період від 1 місяця до 1 року; і з погляду поточного управління, що полягає у зборі та аналізі інформації про відхилення керованого процесу від запланованого стану та наступному розробленню рішень з усунення (мінімізації) цих відхилень.

До теперішнього часу розроблені програмні засоби управління проектами, що забезпечують розв'язання завдань оперативно-календарного планування та управління. Серед найбільш популярних з них програмні засоби Time Line, MS Project (Project Expert), Альт-Інвест, Бізнес-План Про.

Будь-яке управління має на увазі наявність цільової функції, що дозволяє оцінити його результати та ефективність. Ціль оперативно-календарного планування - це мінімізація тривалості виконання проекту при обмеженнях на наявні ресурси. Методи календарного планування були розроблені в 50-х роках майже одночасно: метод критичного шляху та метод оцінки й перегляду програм. Обидва методи складають основу теоретичного апарата методу сіткового планування та управління проектами.

Програмні засоби управління проектами дозволяють:

- побудувати сіткову модель;
- визначити критичний шлях (або кілька критичних шляхів) і його параметри;
- зберігати базовий план проекту та уводити дані про хід робіт;
- виявляти відхилення результатів від базового плану;
- здійснювати управління ресурсами проекту;
- здійснювати управління фінансами проекту.

Сіткове планування та управління включає три основних етапи:

- структурне планування;
- календарне планування;
- оперативне управління.

Структурне сіткове планування передбачає розбивку проекту на чітко визначені операції, для яких визначається тривалість і необхідні ресурси. Потім будується сіткова модель (сітковий графік), що показує взаємозв'язки робіт проекту. Це дозволяє детально аналізувати всі роботи та вносити поліпшення в структуру проекту ще до початку його реалізації. Сітка, що не має числових параметрів, називається структурною. Показник складності мережі обчислюють як відношення кількості робіт до кількості подій.

Календарне сіткове планування передбачає визначення моментів часу початку та закінчення кожної роботи й інші часові характеристики сіткового графіка. Це дозволяє, зокрема, виявляти критичні операції та шляхи сіткової моделі, які визначають термін закінчення проекту. Під час календарного планування визначаються всі часові характеристики всіх робіт і подій з метою наступної оптимізації сіткової моделі, що дозволить поліпшити ефективність використання будь-якого ресурсу (трудових ресурсів, часу, коштів та ін.).

У ході **оперативного сіткового управління** використовуються оптимізований сітковий графік і календарні терміни для складання періодичних звітів про хід виконання проекту. При цьому модель може піддаватися оперативному коректуванню, на підставі якого розробляються нові параметри частини сіткової моделі, що залишилася.

Сіткова модель являє собою план виконання певного комплексу взаємозалежних робіт (операцій), що заданий у формі сіткового графіка, у якому чітко визначені всі часові взаємозв'язки планованих робіт. Математичний апарат сіткових моделей базується на теорії графів.

Графом називається сукупність двох кінцевих множин: множини точок, які називаються **вершинами**, і множини зв'язків між парами вершин, які називаються **ребрами**. Якщо розглянуті пари вершин є впорядкованими, тобто на кожному ребрі задається напрямок, то граф називається орієнтованим; у протилежному випадку – неорієнтованим. Послідовність ребер, що веде від однієї вершини до іншої, утворює **шлях**.

Граф називається **зв'язним**, якщо для будь-яких двох його вершин існує шлях, що їх з'єднує; у протилежному випадку граф називається **незв'язним**.

В економіці та управлінні найчастіше використовується два види графів: дерево та сітка. **Дерево** являє собою зв'язний граф без циклів, що має вихідну вершину (корінь) і крайні вершини; шляхи від вихідної вершини до крайніх вершин називаються гілками. **Сітка** – це орієнтований кінцевий зв'язний граф, що має початкову вершину (джерело) і кінцеву вершину (стік). Таким чином, сіткова модель являє собою граф виду «сітка».

Об’єктом управління в системах СПУ є колективи виконавців, що мають певні ресурси і виконують комплекс операцій, що спрямовані на досягнення наміченої цілі, наприклад, на розробку нової послуги, або на дослідження системи управління підприємством, або на реалізацію комплексу управлінських процедур і операцій для досягнення стратегічної цілі організації та ін.

Приклад сіткової моделі наведений на рисунку 3.1.

Вихідні дані для побудови сіткової моделі, зображеної на рисунку 3.1, подані в таблиці 3.1.

Таблиця 3.1 - Вихідні характеристики робіт

№ з/п	i-j	Q_{ij}	m_{ij}	№ з/п	i-j	Q_{ij}	m_{ij}
1	0-1	20	1	11	5-10	12	3
2	0-2	20	2	12	5-3	16	4
3	0-3	20	3	13	6-11	20	1
4	0-4	14	4	14	7-11	30	7
5	1-5	12	3	15	8-3	0	0
6	1-6	40	2	16	9-12	16	4
7	2-7	0	0	17	10-13	20	5
8	3-7	16	4	18	11-13	20	1
9	4-8	12	3	19	12-14	8	2
10	4-9	6	2	20	13-14	4	1

Позначення в таблиці:

№ п/п - порядковий номер роботи;

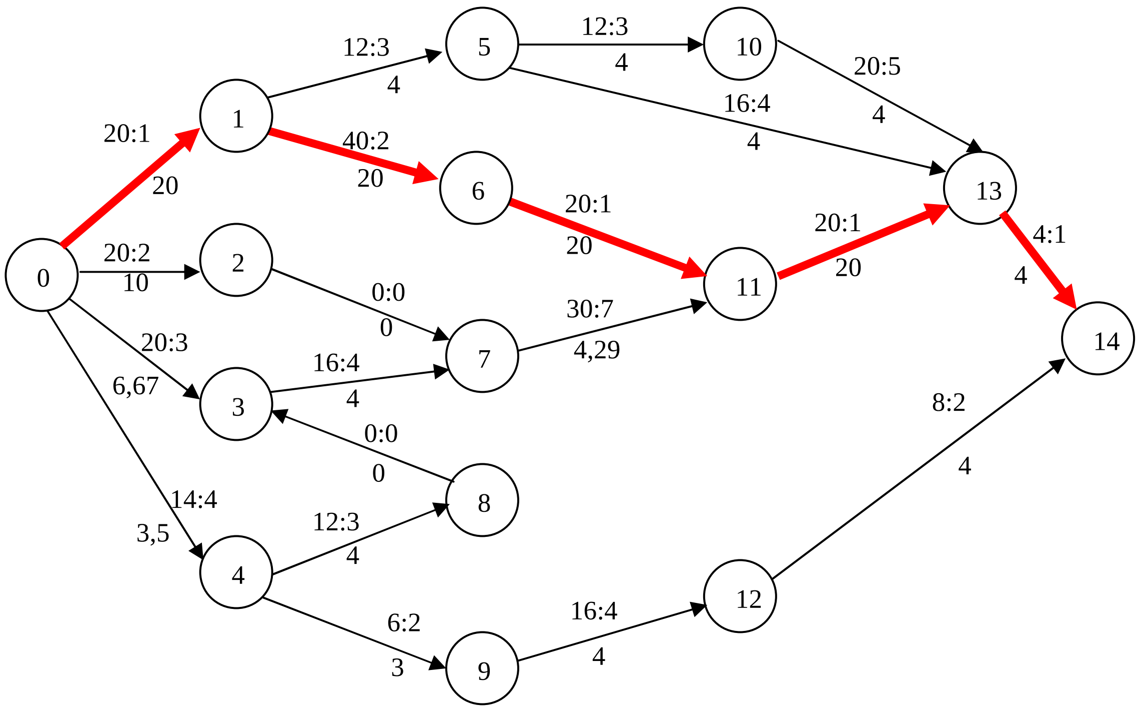
i-j - код роботи, що визначає місце роботи в загальному порядку виконання всього комплексу робіт;

i - попередня щодо роботи подія;

j - наступна за виконаною роботою подія;

Q_{ij} – трудомісткість роботи в людино-годинах або людино-днях;

m_{ij} – кількість виконавців роботи, людин.



Позначення: цифри в кружках – номера подій; стрілки між подіями – роботи; дробові числа над стрілками: чисельник – трудомісткість роботи Q_{ij} у людино-годинах (або людино-днях), а знаменник – кількість виконавців m_{ij} людина; число під стрілкою t_{ij} – тривалість виконання роботи в годинах (або днях).

Рисунок 3.1 - Загальний вигляд сіткової моделі

Елементами сіткової моделі є: роботи, події, шляхи (рис. 3.1).

Робота. У СПУ розрізняють три види робіт:

- будь-який активний трудовий процес, що вимагає витрат часу та ресурсів (енергетичних, трудових, фінансових та ін.);
- очікування - робота, що вимагає витрат часу, але не вимагає ресурсів;
- фіктивна робота - не вимагає ні витрат часу, ні ресурсів. Показує логічний зв'язок між окремими роботами, тобто залежність початку однієї або кількох робіт від інших. Зазвичай дійсні роботи в сітковому графіку позначаються суцільними стрілками, а фіктивні роботи - пунктирними.

Подія – це підсумок зроблених робіт, що дає початок для наступних робіт. Подія не має тривалості в часі. Подія, за якою починається дана робота, називається початковою для даної роботи; вона позначається символом і. Подія, що настає після виконання даної роботи, називається кінцевою для даної роботи; вона позначається символом j. Кожна подія, що включається в сіткову модель, має бути точно й повно визначеною. Її формулювання має містити в собі результат всіх безпосередньо попередніх їй робіт.

У кожній сітці є дві крайніх події – вихідна й завершальна. На рисунку 3.1 це події 0 і 14. **Вихідною** називається подія в сітці, що не має попередніх подій і відбиває початок виконання всього комплексу робіт. Вона позначається символом I. **Завершальною** називається подія, що не має наступних подій і показує досягнення кінцевої цілі виконання всього комплексу робіт. Вона позначається символом J. У ту саму подію може входити й з тієї самої події може виходити кілька видів робіт.

Шлях – це будь-яка послідовність робіт у сітковому графіку, у якій кінцева подія кожної роботи збігається з початковою подією наступної за нею роботи. Якщо відома тривалість кожної роботи t_{ij}, то для кожного шляху можна обчислити його загальний час виконання – довжину, тобто загальну сума тривалостей всіх робіт шляху T_{Li}.

У сітковому графіку треба розрізняти кілька видів шляхів:

- повний шлях L – шлях від вихідної події I до завершальної J; повний шлях з максимальною тривалістю називається **критичним шляхом** $L_{кр}$;
- шлях, що передує події I, $L_{пі}$ – шлях від вихідної події I до даної події i;
- шлях, що іде за подією i, $L_{сі}$ – шлях від даної події і до завершальної події J;
- шлях між подіями і та j;
- підкритичний шлях - повний шлях, найближчий за тривалістю до критичного шляху;
- ненавантажений шлях - повний шлях, тривалість якого значно менша за тривалість критичного шляху.

Сіткову модель будують на початковому етапі планування. Для цього процес реалізації проекту розбивають на окремі роботи, складають перелік робіт і подій, визначають взаємозв'язки робіт, послідовність їхнього виконання і тривалість. Потім будують (зшивають) сітковий графік, для якого розраховують параметри подій і робіт, визначають резерви часу та критичний шлях. Потім проводять аналіз сіткового графіка та його оптимізацію.

Сформулюємо правила побудови сіткової моделі.

Правило 1. Мережа має тільки одну початкову подію й тільки одну кінцеву подію.

Правило 2. Мережу викреслюють ліворуч праворуч. Бажано, щоб кожна подія з більшим порядковим номером зображувалася праворуч щодо попередньої. Для кожної роботи (i-j) має виконуватися загальний напрямок стрілок, що зображують роботи, ліворуч праворуч, при цьому кожна робота повинна виходити з події з меншим номером і входити в подію з більшим номером. Зображення та позначення робіт і подій подані на рисунку 3.2.

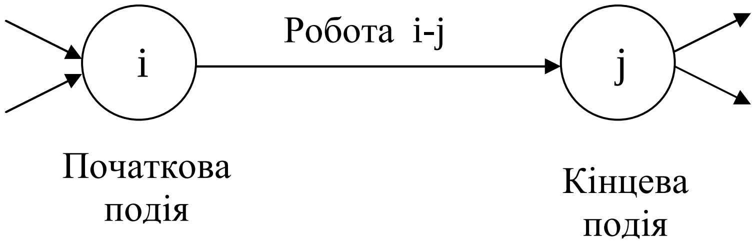


Рисунок 3.2 - Зображення та позначення робіт і подій

Правило 3. Якщо в процесі виконання роботи починається інша робота, що використовує результат певної частини першої роботи, то перша робота розбивається на дві, причому частина першої роботи від початку (0) до видачі проміжного результату, тобто до початку другої роботи, і частина першої роботи, що залишилася, виділяються як дві самостійні роботи.

Правило 4. Якщо n робіт починаються та закінчуються тими самими подіями, то для встановлення взаємно однозначної відповідності між цими роботами і кодами необхідно ввести (n-1) фіктивних робіт. Вони не мають тривалості в часі та уводяться в цьому випадку лише для того, щоб згадані роботи мали різні коди.

Правило 5. У сітці не має бути подій, у які не входить жодної роботи, окрім вихідної події. Порушення цього правила та поява в сітці, окрім вихідної, ще однієї події, у яку не входить жодної роботи, означає або помилку під час побудови сіткового графіка, або відсутність роботи, результат якої необхідний для початку наступної роботи.

Правило 6. У сітці не має бути подій, з яких не виходить жодної роботи, окрім завершальної події. Порушення цього правила та поява в сітці, окрім завершальної, ще однієї події, з якої не виходить жодної роботи, означає або помилку під час побудови сіткового графіка, або планування непотрібної роботи, результат якої нікого не цікавить.

Правило 7. Події потрібно нумерувати так, щоб номер початкової події кожної роботи був менше номера кінцевої події цієї роботи.

Правило 8. У ланцюзі не має бути замкненого контуру.

Побудова сітки є першим кроком на шляху до побудови календарного плану. Другим кроком є розрахунок сіткової моделі, що виконують на сітковому графіку. Для цього користуються простими правилами та формулами, або використовують математичне подання сіткової моделі у вигляді системи рівнянь, цільової функції та граничних умов. Третій крок - оптимізація моделі.

3.4 Часові параметри сіткових моделей

До основних часових параметрів сіткового графіку належать критичний шлях, резерви часу подій і резерви часу робіт.

До власних характеристик робіт належать:

- подвійні індекси робіт ij, що вказують місце роботи в сітковій моделі та взаємозв'язок з іншими роботами і подіями; i - індекс події передуючої початку роботи; j - індекс події, що іде за закінченням роботи;

– трудомісткість роботи в людино-годинах або в людино-днях Q_{ij};

– кількість виконавців m_{ij}, людин;

– тривалість виконання роботи t_{ij} у годинах (або днях) визначається за формулою:

$$t_{ij} = \frac{Q_{ij}}{m_{ij}}. \qquad (3.3)$$

До системних характеристик подій належать: номера (індекси) подій, ранні та пізні терміни настання подій і резерви часу подій.

Номера подій – i або j; система нумерації має забезпечувати умову, щоб для кожної роботи індекси i і j були у співвідношенні i<j.

Ранній термін настання події T_{pi} – це час, який необхідний для виконання всіх робіт, що передують даній події. Він дорівнює найбільшій з тривалостей шляхів, що передують даній події.

Для вихідної події T_{p0} = 0. Для всіх інших подій ранній термін настання визначається за формулами

$$T_{pi} = \max_{Lni}\{TL_{ni}\} \quad \text{або} \quad T_{pi} = \max\left\{\sum_{i=0}^{n} t_{ij}\right\}, \qquad (3.4)$$

де max - максимум береться за всіма роботами (ij) шляхів, що передують події *i*.

Якщо подія *j* має кілька попередніх шляхів, то для визначення раннього терміну здійснення події *j* можна скористатися формулою (рис. 3.3)

$$T_{pi} = \max_{ij}\left[T_{pi} + t(i, j)\right].$$

Пізній термін настання події $T_{пi}$ – це такий час настання події i, перевищення якого спричинить відповідну затримку настання завершальної події сітки. Пізній термін настання будь-якої події дорівнює різниці між тривалістю критичного шляху та найбільшою з тривалостей шляхів, що ідуть за подією *i*. Пізні терміни настання подій розраховуються від поточної до завершальної події. Для завершальної події пізній термін настання дорівнює довжині критичного шляху $T_{пк}$=$T_{Lкр}$, для всіх інших подій (рис. 3.4):

$$T_{пi} = T_{Lкк} - \max_{Lci} T_{Lci} \quad \text{або} \quad T_{пi} = T_{Lкк} - \max\left\{\sum_{i}^{k} t_{ij}\right\}, \qquad (3.5)$$

де $T_{Lкк}$ - тривалість критичного шляху.

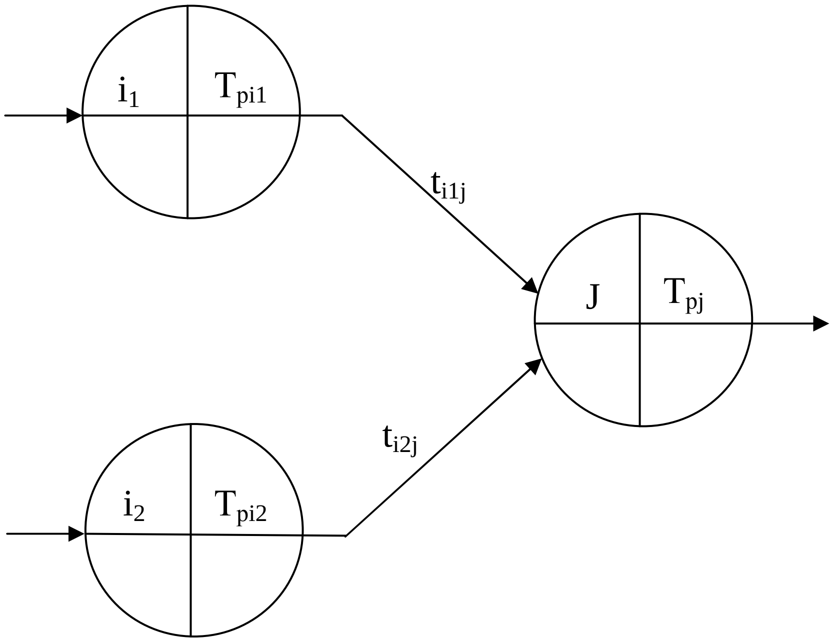


Рисунок 3.3 - Схема розрахунку раннього терміну настання події *i* за формулами (3.4)

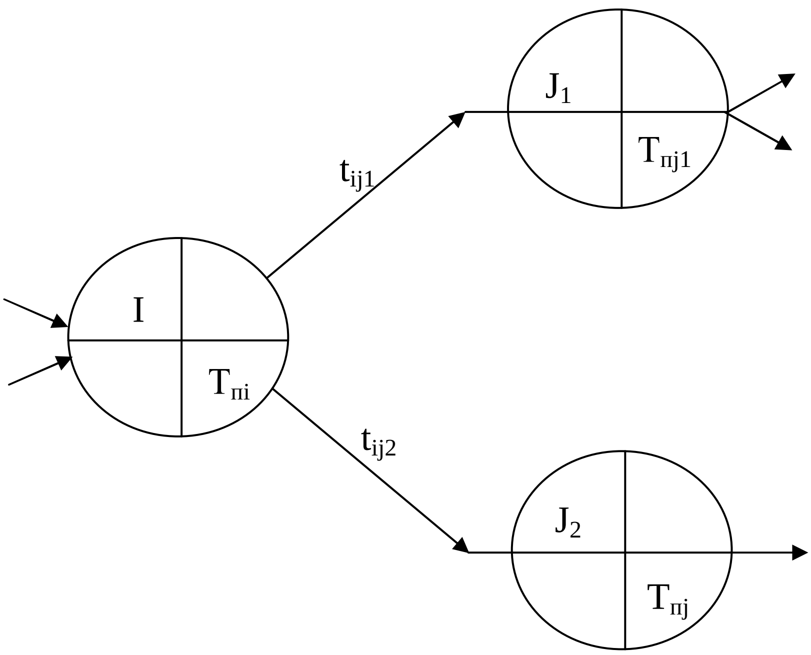


Рисунок 3.4 - Розрахунок пізнього терміну $T_{пi}$ настання події *i* за формулами (3.5)

Якщо подія і має кілька наступних за нею шляхів, то для визначення пізнього терміну здійснення події і можна скористатися формулою (рис. 3.4)

$$T_{пi} = \min_{ij}\left[T_{пj} - t(i, j)\right].$$

Резерв часу настання події і R_i – це такий проміжок часу, на який може бути відстрочене настання події і без порушення термінів завершення проекту в цілому. Початкові та кінцеві події критичних робіт мають нульові резерви часу

$$R_і=T_{пі}-T_{рі}.$$

Розраховані чисельні значення часових параметрів подій іноді записують безпосередньо у вершині сіткового графіка, як представлено на рисунку 3.5.

До системних характеристик робіт належать наступні параметри.

Ранній термін початку роботи $t_{рніj}$ збігається з раннім терміном настання попередньої події *I*, тобто

$$t_{рніj}=T_{рі},$$

тоді ранній термін закінчення роботи $t_{роіj}$ визначається за формулою

$$t_{роіj}=T_{рі} + t_{іj}.$$

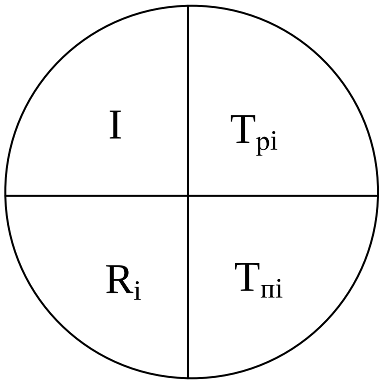


Рисунок 3.5 - Відображення часових параметрів подій у вершинах сіткового графіка

Оскільки робота не може скінчитися пізніше припустимого пізнього терміну своєї кінцевої події *j*, пізній термін її закінчення визначається співвідношенням

$$t_{поіj}=T_{пj},$$

а пізній термін початку роботи $t_{пніj}$ визначається за формулою

$$t_{пніj}=T_{пj} - t_{іj}.$$

Розрізняють чотири види резервів часу робіт.

Повний резерв часу роботи $R_{іj}$ показує, на який час може бути збільшена тривалість $t_{іj}$ окремої роботи іj, щоб при цьому довжина максимального зі шляхів, що проходять через цю роботу, не перевищила б довжини критичного шляху:

$$R_{іj}=T_{пj} - T_{рі} - t_{іj}. \quad (3.6)$$

Вільний резерв часу роботи іj $R_{сіj}$ - це та частина повного резерву, що зберігається в неї за умови, що кінцева подія роботи відбудеться в самий ранній термін, тобто це резерв часу тільки даної роботи, що дозволяє збільшити її тривалість, не викликавши змін ранніх термінів здійснення початкової та кінцевої подій інших робіт:

$$R_{сіj}=T_{рj}-T_{рі}-t_{іj}. \quad (3.7)$$

Вільний резерв часу використовують як допоміжний параметр при необхідності детальнішого аналізу конкретних ситуацій, що виникають у ході виконання робіт.

Частковий резерв часу роботи іj $R_{чіj}$ - це максимальний запас часу, на який можна затримати початок роботи або, якщо вона почалася в ранній термін, збільшити її тривалість, не змінюючи ранніх термінів початку наступних робіт:

$$R_{\text{чij}}=R_{\text{ij}}-R_{\text{сj}} = T_{\text{пj}}-T_{\text{пi}}-t_{\text{ij}}. \tag{3.8}$$

Незалежний резерв часу роботи ij $R_{\text{нij}}$ - це запас часу, при якому всі попередні роботи закінчуються в пізній термін, а всі наступні – починаються в ранній термін. Використання цього резерву не впливає на величину резервів часу інших робіт:

$$R_{\text{нij}}=R_{\text{ij}}-R_{\text{i}}-R_{\text{j}} = T_{\text{рj}}-T_{\text{пi}}-t_{\text{ij}}. \tag{3.9}$$

Частковий резерв часу може бути використаний на збільшення тривалості поточної та наступної робіт без витрат резерву часу попередніх робіт. Вільний резерв часу може бути використаний на збільшення тривалості поточної та попередньої робіт без порушення резерву часу наступних робіт. Незалежний резерв часу може бути використаний для збільшення тривалості тільки поточної роботи.

Роботи, що лежать на критичному шляху, резервів часу не мають. Якщо на критичному шляху лежить початкова подія i, то $R_{\text{ij}} = R_{\text{чij}}$, якщо на критичному шляху лежить кінцева подія j, то $R_{\text{ij}} = R_{\text{сij}}$.

Резерв трудових ресурсів (виконавців робіт) $m_{\text{ij}}\downarrow$:

$$m_{ij} \downarrow = m_{ij} - \frac{Q_{ij}}{t_{ij} + 0{,}5R_{ij}}. \tag{3.10}$$

Розглянемо характеристики шляхів сіткової моделі.

Тривалість шляху є часовою характеристикою й відбиває суму часів на виконання всіх робіт, що належать цьому шляху, між конкретними подіями.

Для повного шляху загальна тривалість робіт дорівнює

$$T_{Ln} = \sum_{i=I}^{i=J} t_{ij}, \tag{3.11}$$

де n - номер повного шляху;

I - початкова подія для всього комплексу робіт;

J - кінцева подія - закінчення робіт усього комплексу.

Для часткового шляху (попереднього події, що іде за подією або між подіями):

$$T_{Lij} = \sum_{i=i\text{н}}^{i=jk} t_{ij}, \tag{3.12}$$

де $i_{\text{н}}$ – подія початку часткового шляху;

$j_{\text{к}}$ – кінцева подія часткового шляху.

Для критичного шляху тривалість шляху є максимальною:

$$T_{L\text{кк}} = \max\left\{\sum_{i=I}^{i=J} t_{ij}\right\}. \tag{3.13}$$

Всі некритичні шляхи сіткового графіка мають резерви часу. Резерв часу шляху R_{L} дорівнює різниці між довжиною критичного і некритичного шляхів

$$R_{\text{L}}=T_{\text{кр}} - T_{\text{L}}.$$

Резерв часу шляху показує, на скільки може бути збільшена сума тривалостей всіх робіт, що лежать на цьому шляху. Звідси випливає, що кожна з робіт некритичного шляху має резерв часу (окрім ділянок, що збігаються з критичним шляхом).

3.5 Аналіз і оптимізація сіткової моделі

Ціль аналізу - виявлення можливостей скорочення термінів розробки в цілому. Аналіз сіткового графіка дозволяє оцінити доцільність структури графіка, завантаження виконавців робіт на всіх етапах виконання проекту, для чого здійснюють групування робіт за величиною резервів та оцінку імовірності завершення проекту в заданий термін. Для визначення напруженості виконання в строк кожної групи робіт використовують коефіцієнт напруженості роботи.

Коефіцієнт напруженості роботи являє собою відношення тривалості незбіжних відрізків шляху, одним з яких є шлях максимальної тривалості, що проходить через роботу ij, а іншим – критичний шлях:

$$K_{нij} = \frac{T_{L\max} - T'_{Lкк}}{T_{Lкк} - T'_{Lкк}}, \text{ або } K_{нij} = 1 - \frac{R_{ij}}{T_{Lкк} - T'_{Lкк}}, \qquad (3.14)$$

де $T'_{Lкк}$ - величина відрізка досліджуваного шляху, що збігається з критичним шляхом;

$T_{Lкк}$ - тривалість критичного шляху;

$T_{L\max}$ - тривалість максимального, але не критичного шляху, що проходить через дану роботу;

R_{ij} – повний резерв часу роботи.

Чим $K_{нij}$ ближче до одиниці, тим складніше виконати дану групу робіт у встановлений термін; чим менше $K_{нij}$, тим більши резерви має цей шлях у сітці. На підставі обчислених значень $K_{нij}$ виділяють три зони: критичну ($K_{нij}$ >0,8), підкритичну (0,6≤$K_{нij}$ ≤0,8) і резервну ($K_{нij}$ <0,6).

В методі PERT важливим завданням аналізу сіткового графіка є оцінка імовірності здійснення завершальної події в заданий термін $p_к$. Для робіт, що носять імовірнісний характер, використовують закон β-розподілу. Числові характеристики β-розподілу $\bar{t}_{ij}$ та σ^2_{ij} для роботи ij розраховують за формулами:

$$\bar{t}_{ij} = \frac{t_{\min ij} + 4t_{нв ij} + t_{\max ij}}{6} \text{ і } \sigma^2_{ij} = \left[\frac{t_{\max ij} - t_{\min ij}}{6}\right]^2, \qquad (3.15)$$

де $t_{\min ij}$ – тривалість роботи ij за самих сприятливих умов (оптимістична оцінка);

$t_{\max ij}$ - тривалість роботи ij за самих несприятливих умов (песимістична оцінка);

$t_{нвij}$ – найбільш імовірна тривалість роботи ij за нормальних умов.

Оскільки оцінка найбільш імовірної тривалості роботи ij фахівцями зазвичай утруднена, у реальних проектах на практиці використовують спрощену формулу середньої тривалості роботи:

$$\bar{t}_{ij} = \frac{2t_{\min} + 3t_{\max}}{5}. \qquad (3.16)$$

Для оцінки імовірності $p_к$ того, що термін виконання проекту $t_{кр}$ не перевершить заданого директивного терміну T, вважають, що $t_{кр}$ є випадковою величиною з нормальним законом розподілу. З статистики відомо, що

$$p_к = P(t_{кр} < T) = 0,5 + 0,5\Phi\left(\frac{T - \bar{t}_{кр}}{\sigma_{кр}}\right),$$

де $\Phi(z) = \Phi\left(\frac{T - \bar{t}_{кр}}{\sigma_{кр}}\right)$ - інтеграл імовірностей Лапласа;

$t_{кр}$ і $\sigma_{кр}$ – середнє значення та середнє квадратичне відхилення довжини критичного шляху, визначаються за формулами (3.15).

Якщо імовірність здійснення завершальної події в заданий термін $p_к$ мала ($p_к < 0,35$), то небезпека зриву заданого терміну велика, і необхідне вживання додаткових заходів. Якщо $p_к>0,65$, то з достатнім ступенем надійності можна прогнозувати виконання робіт проекту у встановлений термін.

Всі обчислення за методом CPM справедливі для методу PERT, якщо замінити значення тривалостей робот їхніми статистичними оцінками.

Пряме завдання оптимізації сіткового графіка: при знайденому критичному шляху використати резерви некритичних робіт і одержати сітку з мінімальними витратами на весь комплекс робіт. Можна поставити й **зворотне завдання**: за рахунок збільшення витрат на роботи критичного шляху скоротити терміни виконання робіт цього шляху, а виходить, і терміни виконання всього комплексу робіт.

Сіткові графіки оптимізують:
- за термінами;
- за використовуваними ресурсами;
- за вартістю.

Для виявлення можливості скорочення часу розробки використовують **метод дослідження ненапружених шляхів**. Цей метод може виконуватися як без урахування, так і з урахуванням впливу на вартість розробки.

Скорочення тривалості критичного шляху може бути досягнуто:
- шляхом перерозподілу ресурсів, як часових, так і трудових або матеріальних та ін., із зон менш напружених у зони напруженіші;
- шляхом зниження трудомісткості критичних робіт за рахунок передачі їхньої частини на інші шляхи, що мають резерви часу;
- шляхом паралельного виконання робіт критичного шляху, а також зміни складу робіт і структури сітки.

В ідеальному випадку довжина повних шляхів може стати рівною довжині критичного шляху, тоді всі роботи вестимуться з рівною напругою, і термін завершення проекту істотно зменшиться.

Під час визначення вартості враховуються використовувані ресурси. Цей метод одержав назву **«час-вартість»**. Він полягає у встановленні залежності між тривалістю та вартістю робіт з метою їхньої оптимізації. Для побудови графіків «час-вартість» для кожної роботи даються:

- нормальна оцінка – мінімально можлива величина грошових витрат C_{min} на виконання роботи (при цих витратах робота може бути виконана за максимальний час T_{max});

- мінімальна оцінка - мінімально можливий час виконання роботи T_{min}; цьому часу відповідатимуть підвищені розміри грошових витрат C_{max}, на виконання роботи.

Графік за допомогою апроксимуючої прямої (рис. 4.6) дозволяє визначити розміри збільшення витрат при необхідності скорочення терміну виконання роботи, або при розв'язанні зворотного завдання – збільшення часу виконання роботи, якщо необхідно витрати зменшити. Шукана величина витрат ΔC, що необхідні для виконання роботи в скорочений час $T_{иск}$, дорівнює:

$$\Delta C = \frac{(C_{\max} - C_{\min})(t_{\max} - t_{ij})}{(t_{\max} - t_{\min})} \text{ грн.},$$

де $\frac{(C_{\max} - C_{\min})}{(t_{\max} - t_{\min})}$ - тангенс кута α, що чисельно дорівнює коефіцієнту зростання витрат K_s на одиницю часу, грн/од. часу.

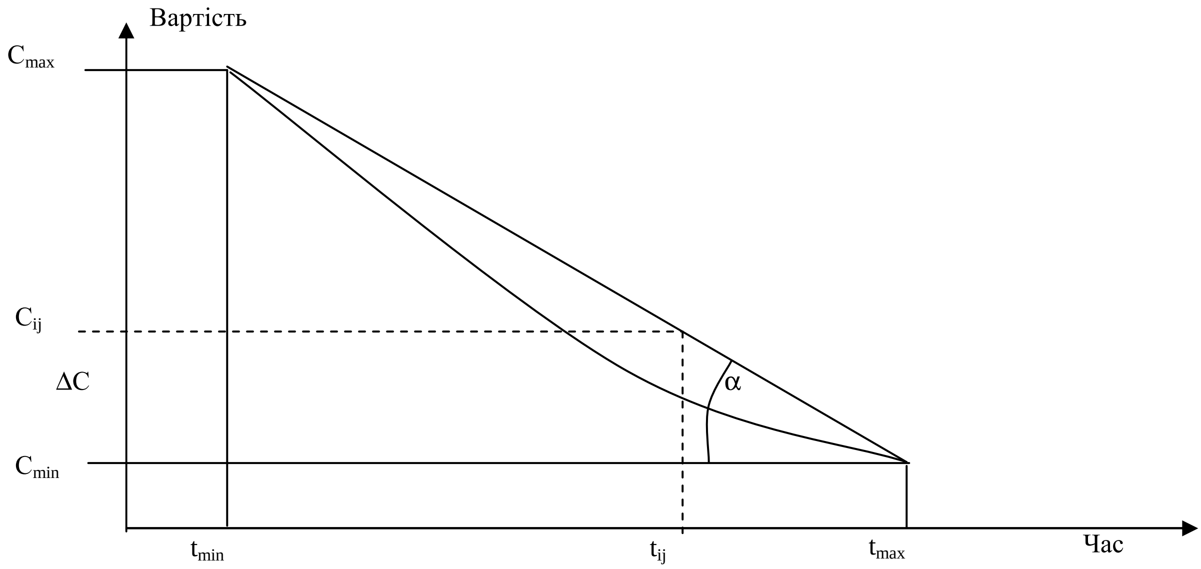


Рисунок 3.6 - Залежність «час-вартість»

Оптимізація сіткового графіка має метою скорочення довжини критичного шляху, вирівнювання коефіцієнтів напруженості робіт, раціонального використання ресурсів.

Таким чином, організаційна процедура сіткового планування передбачає виконання наступних етапів:

- розробка планів виконання всіх окремих робіт комплексу; визначення їхньої трудомісткості і необхідних трудових та інших ресурсів;
- встановлення взаємозв'язків робіт та їхнього відносного порядку виконання;
- формування подій і привласнення їм номерів за розглянутими раніше правилами. Результати виконання етапів оформляються у вигляді таблиці (наприклад, таблиці 3.1);
- складання сіткового графіку (моделі) за відповідними правилами;

- розрахунок всіх параметрів сітки за формулами і визначення критичного шляху;

- виконання аналізу сітки. За коефіцієнтами напруженості та резервами часу класифікують шляхи та роботи;

- виконання оптимізації сітки - часткової або комплексної за одним з критеріїв.

Під час використання часткового критерію «мінімум часу виконання всього комплексу робіт» визначається можливість скорочення тривалості робіт критичного шляху за рахунок ненапружених шляхів - перерозподіл ресурсів (переведення частини працівників на роботи критичного шляху), тобто проводиться варіювання кількості працівників для робіт ненапружених і робіт критичного шляху.

Під час використання критерію «мінімум витрат» підраховуються витрати на заробітну плату працівників, що зайняті виконанням комплексу робіт. Перелік фахівців, їхні оклади, час роботи, кількість працівників кожної кваліфікації зводяться в таблицю.

Управління ходом робіт за допомогою сіткового графіка починається після того як вихідний план затверджений, доведений до всіх відповідальних виконавців і починаються перші роботи, що спираються на вихідну подію. Закінчується процес управління ходом робіт у момент здійснення завершальної події.

Періодично повторювані етапи оперативного управління включають:

- збір проміжних звітів про хід робіт від відповідальних виконавців;

- обробку отриманої інформації та аналіз змін, що внесені відповідальними виконавцями стосовно первісних оцінок;

- підготовку рішень і перевірку їх за допомогою розрахунків нового сіткового графіка вручну або на ЕОМ з метою зведення до мінімуму розбіжностей і оптимізації плану робіт;

- прийняття остаточних рішень, перебудова вихідного графіка, розробка календарних планів-графіків для відповідальних виконавців;

- вироблення вихідної інформації та доведення її до всіх рівнів керівництва, у т.ч. до відповідальних виконавців.

До переваг СПУ відносять наступне:

- увага керівництва концентрується на вирішальних роботах;

- у будь-який момент керівництво має вичерпну інформацію;

- реалізується принцип безперервності планування ходу робіт і управління ними;

- система забезпечує можливість раціонального маневрування виділеними для даної розробки ресурсами;

- встановлюється чіткий взаємозв'язок між відповідальними виконавцями окремих робіт.

Розглянемо приклад побудови та оптимізації сіткової моделі.

Визначення вихідної тривалості робіт зробимо за формулою (3.3)

$$t_{ij}=Q_{ij}/m_{ij}.$$

Результати розрахунків у годинах наступні

$$t_{0-1}=20, \qquad t_{5-10}=4,$$

	t_{5-13}=4,
t_{0-3}=6,67,	t_{6-11}=20,
	t_{7-11}=4,29,
t_{1-5}=4,	t_{8-13}=0,
t_{1-6}=20,	t_{9-12}=6,
t_{2-7}=0,	t_{10-13}=4,
	t_{11-13}=20,
t_{4-8}=4,	t_{12-14}=4,
t_{4-9}=3,	t_{13-14}=4.

Побудова вихідної сіткової моделі. Використовуючи 8 правило, побудуємо сіткову модель за вихідним даними графічним способом (рис. 3.1).

Визначення та аналіз системних характеристик вихідної сіткової моделі. Визначимо всі можливі повні шляхи сіткової моделі (рис. 3.1), які подамо як ланцюжки подій від початкової до кінцевої події:

Шлях L_1:0 - 1 - 5 - 10 - 13 – 14,
Шлях L_2:0 - 1 - 5 - 13 – 14,
Шлях L_3:0 - 1 - 6 - 11 - 13 – 14,
Шлях L_4:0 - 2 - 7 - 11 - 13 – 14,
Шлях L_5:0 - 3 - 7 - 11 - 13 – 14,
Шлях L_6:0 - 4 - 8 - 3 - 7 - 11 - 13 – 14,
Шлях L_7:0 - 4 - 9 - 12 - 14.

Можливих шляхів сім. Зробимо розрахунки, за допомогою яких обчислимо тривалості кожного шляху. Для цього скористаємося формулою (3.11)

$$T_{Li} = \sum_{i=I}^{i=J} t_{i=j},$$

де t_{i-j} – тривалості робіт даного шляху (у годинах).

T_{L1} = 20 + 4 + 4 + 4 + 4 = 36 годин,

T_{L2} = 20 + 4 + 4 + 4 = 32 години,

T_{L3} = 20 + 20 + 20 + 20 + 4 = 84 години,

T_{L4} = 10 + 0 + 4,29 + 20 + 4 = 38,29 годин,

T_{L5} = 6,67 + 4 + 4,29 + 20 + 4 = 38,96 годин,

T_{L6} = 3,5 + 4 + 0 + 4 + 4,29 + 20 + 4 = 39,79 годин,

T_{L7} = 3,5 + 3 + 4 + 4 = 14,5 годин.

Виділимо критичний шлях $L_{кр}$. Шлях з найбільшою тривалістю за часом буде критичним. Це шлях L_3 з тривалістю T_{L3} = 84 години.

Розрахуємо резерв часу R_{Li} для кожного шляху L_i. Резерв часу обчислюється за формулою $R_L = T_{Lкр}-T_{Li}$. Дані про тривалість шляхів і резерви часу за шляхами наведені в таблиці 3.2.

Розрахуємо характеристики подій. При визначенні ранніх термінів настання подій $T_{рi}$ рухаємося за графіком ліворуч праворуч і використовуємо формулу (3.4). При визначенні пізніх термінів настання подій $T_{пi}$ рухаємося за графіком праворуч ліворуч і використовуємо формули (3.5).

Таблиця 3.2 - Вихідні тривалості та резерви шляху

Шлях *L*i	T_L**, год**	R_L**, год**
1	36	48,0
2	32	52,0
3	84	0,0
4	38,29	45,7
5	38,96	45,0
6	39,79	44,2
7	14,5	69,5

Обчислені характеристики подій представлені в таблиці 3.3.

Таблиця 3.3 - Ранні та пізні терміни настання подій

Подія L_t	T_{Pi}	$T_{\text{пi}}$
0	0,0	0,0
1	20,0	20,0
2	10,0	55,7
3	6,7	43,7
4	3,5	47,7
5	24,0	72,0
6	40,0	40,0
7	11,5	55,7
8	7,5	43,7
9	6,5	76,0
10	28,0	76,0
11	60,0	60,0
12	10,5	80,0
13	80,0	80,0
14	84,0	84,0

Обчислимо максимальний запас часу, на який можна відстрочити початок або збільшити тривалість кожної з робіт без збільшення тривалості критичного шляху. Цей запас називається резервом часу роботи і позначається R_i. Для цього скористаємося формулою (3.6).

Роботи на критичному шляху не мають повного резерву часу, для них $R_{ij} = 0$, тоді одержуємо, що

$R_{0\text{-}1}$=0, $R_{5\text{-}10}$=48,
$R_{0\text{-}2}$=45,71, $R_{5\text{-}13}$=52,
$R_{0\text{-}3}$=45,04, $R_{6\text{-}11}$=0,
$R_{0\text{-}4}$=44,21, $R_{7\text{-}11}$=44,21,
$R_{1\text{-}5}$=48, $R_{8\text{-}3}$=44,21,
$R_{1\text{-}6}$=0, $R_{9\text{-}12}$=73,5,
$R_{2\text{-}7}$=45,71, $R_{10\text{-}13}$=48,
$R_{3\text{-}7}$=44,21, $R_{11\text{-}13}$=0,
$R_{4\text{-}8}$=44,21, $R_{12\text{-}14}$=69,5,
$R_{4\text{-}9}$=69,5, $R_{13\text{-}14}$=0.

Оптимізація сіткової моделі (рис. 3.1) за критерієм «мінімум часу» дозволяє вирішити завдання скорочення часу виконання всього комплексу робіт. Мінімізація часу виконання всього проекту можлива тільки за рахунок скорочення тривалості виконання робіт, що лежать на критичному шляху $L_К$. Для цього з ненавантажених шляхів ми маємо зняти виконавців і рівномірно розподілити їх по критичному шляху та близьким до критичного (підкритичним) шляхам.

Введемо нові позначення: $m_{i\_j}\downarrow$ - виконавці, що зняті з роботи i-j; $m_{i\_j}\uparrow$ - виконавці, що призначені на роботу i-j; m'_{i_j} - нова кількість виконавців на роботі i-j; t_{ij} - нова тривалість роботи i-j.

Кількість виконавців $m_{l\text{-}j}\downarrow$, яких можливо зняти з робіт обчислимо за формулою (3.10) $m_{ij}\downarrow = m_{ij} - \dfrac{Q_{ij}}{t_{ij} + 0{,}5R_{ij}}$. Під час перерозподілу виконавців необхідно дотримувати умов:

а) $m'_{i\text{-}j} = m_{i\text{-}j} + m_{i\text{-}j}\uparrow$ - для критичного шляху;

б) $m'_{i\text{-}j} = m_{i\text{-}j} + m_{i\text{-}j}\downarrow$ - для ненавантажених шляхів;

в) $\sum m_{i\text{-}j}\downarrow = \sum m_{i\text{-}j}\uparrow$ - загальна сума виконавців знятих з робіт ненавантажених і доданих на роботи критичного шляху.

$m_{0\text{-}1}\downarrow=0$, $m_{5\text{-}10}\downarrow=2$,
$m_{0\text{-}2}\downarrow=1$, $m_{5\text{-}13}\downarrow=3$,
$m_{0\text{-}3}\downarrow=2$, $m_{6\text{-}11}\downarrow=0$,
$m_{1\text{-}5}\downarrow=2$, $m_{8\text{-}3}\downarrow=0$,
$m_{1\text{-}6}\downarrow=0$, $m_{9\text{-}12}\downarrow=3$,
$m_{2\text{-}7}\downarrow=0$, $m_{10\text{-}13}\downarrow=4$,
$m_{3\text{-}7}\downarrow=3$, $m_{11\text{-}13}\downarrow=0$,
$m_{4\text{-}8}\downarrow=2$, $m_{12\text{-}14}\downarrow=1$,
$m_{4\text{-}9}\downarrow=1$, $m_{13\text{-}14}\downarrow=0$.

З можливих варіантів $m_{i\text{-}j}\downarrow$ оберемо роботи i-j, з яких найзручніше зняти виконавців. Для цього ми проведемо власне оптимізацію даного проекту безмашинним способом, переставляючи виконавців з ненавантажених шляхів L_i на роботи i-j критичного шляху $L_{кр}$. Перестановки виконавців і результати оптимізації відбиті в таблиці 3.4.

Таблиця 3.4 - Результати перерозподілу трудових ресурсів (виконавців)

i-j	**$Q_{i\text{-}j}$**	**$m_{i\text{-}j}$**	**$t_{i\text{-}j}$**	**$m_{i\text{-}j}\downarrow$**	**$m_{i\text{-}j}\uparrow$**	**$m'_{i\text{-}j}$**	**$t'_{i\text{-}j}$**
0-1	20	1	20		3	4	5
0-2	20	2	10				
0-3	20	3	6,67				
0-4	14	4	3,5	1		3	4,67
1-5	12	3	4	1		2	6
1-6	40	2	20		3	5	8
2-7	0	0	0				
3-7	16	4	4				
4-8	12	3	4				

Продовження таблиці 3.4

i-j	Q_{i-j}	m_{i-j}	t_{i-j}	$m_{i-j}\downarrow$	$m_{i-j}\uparrow$	m'_{i-j}	t'_{i-j}
4-9	6	2	3	1		1	6
5-10	12	3	4	1		2	6
5-13	16	4	4	2		2	8
6-11	20	1	20		3	4	5
7-11	30	7	4,29	2		5	6
8-3	0	0	0				
9-12	16	4	4	2		2	8
10-13	20	5	4	1		4	5
11-13	20	1	20		3	4	5
12-14	8	2	4	1		1	8
13-14	4	1	4				

Визначимо нову тривалість часу виконання всіх робіт кожного шляху після оптимізації

T_{L1} =5 + 6 + 6 + 5 + 4 = 26 (годин),
T_{L2} = 5 + 6 + 8 + 4 = 23 (годин),
T_{L3} =5 + 8 + 5 + 5 + 4- 21 (годин),
T_{L4} =10 + 0 + 6 + 5 + 4 = 25 (годин),
T_{L5} = 6,67 + 4 + 6 + 5 + 4 = 25,67 (годин),
T_{L6} = 4,67 + 4 + 0 + 4 + 6 + 5 + 4 = 27,67 (годин),
T_{7i} = 4,67 + 6 + 8 + 8 = 26,67 (годин).

Для зручності порівняння тривалості шляхів вихідної сіткової моделі з тривалістю шляхів оптимізованої сіткової моделі звернемося до таблиці 3.5.

Таблиця 3.5 - Системні характеристики вихідної та оптимізованої сіткової моделі

Шлях L_i	T_{Li} год	T'_{Li} год
1	36	39,3
2	32	25,3
3	84	15,1
4	38,29	44,9
5	38,96	42,5
6	38,79	30,3
7	15,5	37,0
$\sum$	283,5	234,5
L_{cp}	40,65	33,5

Загальний час виконання комплексу робіт скоротився.

Контрольні запитання

1. Охарактеризуйте зміст функцій координації та регулювання як функцій управління.

2. У яких ситуаціях виникають задачі упорядкування та координації? Чим

вони відрізняються від завдань масового обслуговування?

3. Поясніть, які властивості методу динамічного програмування дозволяють використовувати його для оптимізації задач упорядкування перебору варіантів.

4. Чим схожі завдання комбінаторної оптимізації та дискретного програмування?

5. Чи використовують методи дискретного програмування для розв'язання задач комбінаторної оптимізації?

6. Поясніть, чи є комбінаторним алгоритмічний метод віток і границь і чому?

7. Назвіть та охарактеризуйте часткові завдання комбінаторної оптимізації.

8. Сформулюйте задачу про ранець та задачу про комівояжера.

9. Сформулюйте задачу цілочислового програмування у загальному вигляді.

10. Сформулюйте завдання двопроцесорного обслуговування Джонсона.

11. Охарактеризуйте метод сіткового планування та управління. Які завдання він вирішує?

12. Охарактеризуйте методи CPM та PERT. Чим вони відрізняються?

13. Охарактеризуйте основні можливості програмних засобів управління проектами.

14. Назвіть та охарактеризуйте основні етапи сіткового планування та управління.

15. Поясніть сутність елементів сіткової моделі - робот, подій, шляхів.

16. Які основні правила побудови сіткової моделі Ви знаєте?

17. Охарактеризуйте основні часові параметри сіткового графіку.

18. Які параметри належать до характеристик робіт сіткового графіку?

19. Які параметри належать до характеристик подій сіткового графіку?

20. Поясніть, що являють собою ранній та пізній терміни настання події?

21. Поясніть зміст повного та вільного, часткового та незалежного резервів часу.

22. Поясніть, у чому полягає аналіз сіткового графіку?

23. З якою метою обчислюють коефіцієнт напруженості роботи?

24. Як оцінюють імовірність, що термін виконання проекту не перевершить заданого директивного терміну?

25. За якими критеріями оптимізують сіткові графіки?

26. Які заходи дозволяють скоротити тривалість критичного шляху?

27. Поясніть сутність методу «час-вартість».

28. Які етапи містить організаційна процедура сіткового планування?

29. Які етапи містить оперативне управління графіком робот?

30. Охарактеризуйте переваги та недоліки сіткового планування та управління.

ТЕМА 4
ЗАДАЧІ ТА МОДЕЛІ ЗАМІНИ

4.1 Сутність та класифікація задач заміни

Сутність завдання заміни обладнання полягає в наступному. У процесі експлуатації обладнання перетерплює моральне та фізичне зношування. У результаті цього знижується його продуктивність та ліквідаційна вартість, збільшуються витрати на його обслуговування та ремонт. Таким чином, перед підприємством постає завдання про доцільність продовження експлуатації обладнання.

Залежно від характеру обладнання процес заміни включає два класи. Перший пов'язаний з обладнанням, що призначено для тривалої експлуатації. Воно за ступенем старіння стає менш продуктивним фізично внаслідок зношення або морально в результаті появи нових, досконаліших технологій (сюди відносяться, наприклад, металорізальні верстати, автомобілі та ін.). Експлуатація застарілого обладнання пов'язана з збільшенням виробничих витрат, подовженням часу простою, збільшенням числа відмов і тривалості ремонту та ін. Одночасно заміна старого обладнання новим також сполучена з витратами. Необхідно визначити такий термін служби обладнання, при якому дохід від придбаного нового обладнання починає перевищувати первісні вкладення на його придбання.

При оренді обладнання необхідно враховувати, що із зростанням терміну оренди зменшується орендна плата в одиницю часу, але при цьому зростають експлуатаційні витрати.

Завдання першого класу належать до класу задач динамічного програмування. Під час побудови моделі задачі про заміну обладнання як критерій оптимальності зазвичай приймають прибуток від його експлуатації (задача максимізації) або витрати на його обслуговування (задача мінімізації). При цьому вважають, що питання про заміну обладнання вимагає рішення на початку кожного проміжку експлуатації, наприклад, на початку кожного року, і що термін експлуатації обладнання не обмежений. Зокрема, на початку кожного року потрібно прийняти одне з трьох рішень:

- продати обладнання та замінити його новим;
- ремонтувати обладнання та продовжувати його експлуатацію;
- продовжувати експлуатацію обладнання без ремонту.

Другий клас задач пов'язаний з обладнанням, що має малий термін служби, наприклад, лампи освітлення, елементи мікросхем та ін. Під час розв'язання задач другого класу доводиться визначати, які саме одиниці обладнання треба замінити і як часто потрібно робити заміну для того, щоб мінімізувати загальні витрати. Якщо заміну обладнання робити лише після його виходу з ладу, то при мінімумі витрат на відновлення зростають витрати, що пов'язані з простоями, тоді як заміна деталей до їхньої поломки підвищує вартість обладнання. Базою для розв'язання цих задач є теорія надійності, зокрема, такий показник надійності як частота відмов залежно від терміну служби обладнання.

На практиці для визначення періоду експлуатації обладнання і, після якого має бути зроблена його заміна, використовують статистичний метод, що заснований на наступних міркуваннях:

- якщо експлуатаційні витрати c(t) у наступному періоді нижче середньої величини витрат у попередніх періодах, то обладнання продовжують експлуатувати,

$$c_i < \frac{c_1 + c_2 + ... + c_i}{i} \quad ; \qquad (4.1)$$

- якщо експлуатаційні витрати в наступному періоді перевершують величину середніх витрат у попередніх періодах, то обладнання треба замінити,

$$c_i > \frac{c_1 + c_2 + ... + c_i}{i} \quad . \qquad (4.2)$$

Приклад 4.1. Витрати, що пов'язані з придбанням і заміною обладнання, подані в таблиці 4.1. Визначити, коли потрібна заміна обладнання.

Таблиця 4.1 – Вихідні дані

Період (рік експлуатації)	**Фактичні витрати, грн.**	**Середні витрати, грн.**
1	50	50
2	10	30
3	20	26,7
4	30	27,5
5	40	30
6	50	33,3

З розрахунків видно, що фактичні витрати на експлуатацію обладнання перевищуватимуть середнє значення витрат за попередній період у четвертому періоді експлуатації, звідки випливає, що заміну обладнання доцільно зробити після трьох періодів експлуатації

$$c_4 = 30 > \frac{50 + 10 + 20 + 30}{4} = 27{,}5 \quad .$$

4.2 Загальна схема методу динамічного програмування

Метод динамічного програмування призначений для багатоетапних (багатокрокових) операцій. Показник ефективності *W* може бути адитивним, тобто визначатися сумою показників ефективності w_i на окремих етапах

$$W = \sum_{i=1}^{n} w_i \quad , \qquad (4.3)$$

або мультиплікативним, тобто визначатися добутком показників ефективності w_i

$$W = \prod_{i=1}^{n} w_i \quad . \qquad (4.4)$$

На кожному кроці вибирається розв'язок x_i, що називають **умовним кроковим управлінням**. Сукупність таких розв'язків є управлінням операцією в цілому *x*

$$x = (x_1, x_2, ..., x_n) \quad (4.5)$$

Потрібно знайти таке управління x^*, при якому показник ефективності W обертається на максимум:

$$W = \sum_{i=1}^{n} w_i \Rightarrow \max \quad (4.6)$$

Управління x^*, при якому W обертається на максимум, називається **оптимальним управлінням**

$$x^* = (x^*_1, x^*_2, ..., x^*_n).$$

В основі методу динамічного програмування лежить покрокова оптимізація, зміст якої збігається до визначення оптимального управління на кожному кроці з урахуванням його майбутніх наслідків на майбутніх кроках. Інакше кажучи, на кожному кроці шукається таке управління, що забезпечує оптимальне продовження процесу щодо досягнутого в цей момент стану. Цей принцип вибору управління називається **принципом оптимальності**. Сформулюємо його:

> **Який би не був стан системи *S* перед черговим кроком, треба вибирати управління на цьому кроці так, щоб виграш на даному кроці плюс оптимальний виграш на всіх наступних кроках був максимальним.**

Таким чином, управління на і-му кроці обирається так, щоб була максимальною сума показників ефективності на всіх кроках, що залишилися до кінця процесу управління, плюс даний. Очевидно, що найпростіше вибирати оптимальне управління x_n на останньому n-му кроці, тому процес динамічного програмування розвертають від кінця до початку. При цьому на останньому n-му кроці розглядають кілька управлінь, виходячи з усієї сукупності припущень про те, як закінчився передостанній (n-1)-й крок. Ці управління називають **умовними оптимальними управліннями**. Потім оптимізують управління на передостанньому (n-1)-му кроці, базуючись на припущеннях про те, чим закінчився (n-2)-й крок. Так, просуваючись назад, доходять до першого кроку, після чого можна побудувати вже не умовне, а просто оптимальне управління x^* і визначити оптимум критерію ефективності W^*.

Нехай на початку першого кроку об'єкт управління знаходився у стані S_0. Уживши оптимальне управління x^*_1 на першому кроці, ми змінимо стан об'єкта на певний новий S^*_1, потім, уживши відоме умовне оптимальне управління x^*_2 на другому кроці, ми переведемо об'єкт у стан S^*_2, і так далі.

Таким чином, у процесі оптимізації багатокроковий процес проходять двічі, перший раз - від кінця до початку, а другий - від початку до кінця.

Задача покрокової оптимізації формулюється в такий спосіб:

> знайти таке припустиме управління x^*, що переводить систему зі стану S_0 у стан S_n, при якому цільова функція $W=f(S_0,x)$ приймає екстремальне значення (мінімум або максимум).

Відзначимо особливості моделі динамічного програмування.

1. Задача оптимізації інтерпретується як n-кроковий процес управління.

2. Цільова функція дорівнює сумі цільових функцій n кроків, тобто є адитивною

$$W = \sum_{i=1}^{n} f_i(S_{i-1}, x_i) \quad , \qquad (4.7)$$

де $f_i(S_{i-1}, x_i) = w_i$ - цільова функція і-го кроку.

3. Вибір управління на і-му кроці залежить тільки від стану системи S_{i-1} і не впливає на попередні кроки, тобто зворотний зв'язок відсутній.

4. Стан S_i системи наприкінці і-го кроку залежить тільки від попереднього стану S_{i-1} і управління на і-му кроці x_i і не залежить від попередніх станів і управлінь, тобто відсутня післядія

$$S_i = \varphi_i(S_{i-1}, x_i), \quad i = \overline{1, n} \; . \qquad (4.8)$$

Даний вираз називається рівняннями станів.

5. На кожному кроці управління x_i залежить від кінцевого числа керуючих змінних, а стан S_i - від кінцевого числа параметрів.

Сформулюємо ряд практичних рекомендацій щодо постановки задач динамічного програмування, сформувавши порядок дій.

1. Вибрати параметри (фазові координати), що характеризують стан керованої системи S перед кожним кроком.

2. Розділити операцію управління на етапи (кроки).

3. Визначити набір крокових управлінь x_i для кожного кроку та обмеження, що накладаються на них.

4. Записати функцію цілі для і-го кроку

$$w_i = f_i(S_{i-1}, x_i)$$

і обчислити значення показника ефективності на і-му кроці при управлінні x_i, якщо перед цим система була у стані S_{i-1}.

5. Записати функції зміни стану системи

$$S_i = \varphi_i(S_{i-1}, x_i), \quad i = \overline{1, n}$$

і визначити, як зміниться стан S на і-му кроці при управлінні x_i.

6. Записати основне рекурентне рівняння динамічного програмування, що виражає умовний оптимум $w_i(S)$, починаючи з і-го кроку

$$w_i(S) = \max_{xi}\{f_i(S_i, x_i) + w_{i+1}(\varphi i(S_{i+1}, x_i))\} \quad , \qquad (4.9)$$

де x_i – умовне оптимальне управління на і-му кроці;

$w_i(S) = w_{i+1}(\varphi i(S_{i+1}, x_i))$ - уже відома функція цілі наступних кроків, у її вираз треба підставити змінений стан S.

7. Зробити умовну оптимізацію останнього n-го кроку, задаючись рядом станів S, з яких можна за один крок дійти до кінцевого стану, обчислюючи для кожного з них умовний оптимум цільової функції за формулою:

$$w_n(S) = \max_{xn}\{f_n(S_n, x_n)\} \quad , \qquad (4.10)$$

і визначити умовне оптимальне управління $x_n(S)$, для якого цей оптимум досягається.

8. Зробити умовну оптимізацію (n-1)-го, (n-2)-го та інших кроків за рекурентною формулою (5.9), покладаючи у ній і=(n-1), (n-2),…, і для кожного з кроків відзначити умовне оптимальне управління $x_i(S)$, при якому досягається оптимум.

Слід зазначити, що стан системи в початковий момент S_0, як правило, відомий. Тому на останньому кроці зворотного прогону оптимум цільової функції є оптимумом всієї операції

$$W^*=w_1(S_0).$$

9. Зробити безумовну оптимізацію управління, для чого взяти знайдене оптимальне управління на першому кроці $x^*_1=x_1(S_0)$ і змінити стан системи за формулою (4.8). Для знову знайденого стану знайти оптимальне управління на другому кроці x^*_2, і так до кінця.

4.3 Динамічна модель задачі про заміну обладнання

Задача заміни обладнання збігається до визначення оптимального терміну його експлуатації. При цьому критерієм ефективності може бути прибуток від експлуатації обладнання (задача максимізації прибутку) або витрати на обслуговування обладнання (задача мінімізації витрат).

Елементи моделі динамічного програмування визначаються в такий спосіб. Етап і подається порядковим номером року і, $i=\overline{1,n}$. Параметром стану є вік обладнання. На початку першого року експлуатації обладнання нове й $S_0=0$, у наступні роки $S_i=t$. Варіантами розв'язання на і-му етапі є альтернативи: продовжити експлуатацію або замінити механізм на початку і-го року. Можливе управління на кожному кроці має якісні ознаки: зберегти експлуатацію обладнання без ремонту $x^{с}$, відремонтувати обладнання та продовжувати його експлуатацію $x^{р}$ і замінити обладнання новим $x^{з}$.

На початку кожного року приймається рішення або про експлуатацію обладнання ще один рік, або про заміну його новим. Позначимо $r(t)$ прибуток від експлуатації t-літнього механізму, $c(t)$ - витрати на його обслуговування протягом року, $p(t)$ - вартість продажу обладнання, що експлуатувалося t років. Вартість нового обладнання залишається незмінною протягом всіх n років і дорівнює P.

У задачі максимізації прибутку від експлуатації обладнання критерій ефективності w_i, якому на і-му кроці відповідає умовне оптимальне управління $x_i(S)$, обчислюється на підставі рекурентного виразу

$$w_i(t)=\max\left\{\begin{array}{ll} r(t)-c(t)+w_{i+1}(t+1), & \textit{якщо експлуатувати обладнання,} \\ r(0)+p(t)-P-c(0)+w_{i+1}(1), & \textit{якщо замінити обладнання.}\end{array}\right\}. \qquad (4.11)$$

У задачі мінімізації витрат на експлуатацію обладнання критерій ефективності w_i обчислюється на підставі наступного рекурентного виразу:

$$w_i(t)=\min\left\{\begin{array}{ll} c(t)+w_{i+1}(t+1), & \textit{якщо експлуатувати обладнання,} \\ P+c(0)-p(t)+w_{i+1}(1), & \textit{якщо замінити обладнання}\end{array}\right\}. \qquad (4.12)$$

Рівняння (4.11) і (4.12) показують, що функція цілі $w_i(t)$ на і-му етапі повинна бути виражена як функція наступного етапу.

У термінології динамічного програмування S_i називається ***станом системи*** на етапі і. Стан системи на етапі і – це інформація, що зв'язує етапи один з одним, при цьому оптимальні рішення для етапів, що залишилися, можуть прийматися без повторної перевірки того, як були отримані рішення на попередніх етапах. Таке визначення стану системи дозволяє розглядати кожний етап окремо й гарантує, що рішення є припустимим на кожному етапі. Таким чином, можна сформулювати ***принцип оптимальності***:

> **На кожному етапі оптимальна стратегія визначається незалежно від стратегій, що використовані на попередніх етапах. (Оптимальна стратегія управління не залежить від передісторії системи, а залежить від стану системи в сучасний момент часу й цілі управління).**

Приклад 4.2. Нехай є обладнання, що експлуатується протягом 5 років, а потім продається. Вартість нового обладнання P=4000 грн. Ліквідаційна вартість обладнання залежить від часу експлуатації $p(t)=P*2^{-t}$. Витрати на утримання обладнання також залежать від часу його експлуатації c(t)=600(t+1). Визначити оптимальну стратегію експлуатації обладнання протягом 5 років, при якій сумарні витрати з урахуванням початкової покупки та заключного продажу будуть мінімальними.

Розв'язання

Термін експлуатації 5 років, причому, по закінченні 5 років обладнання продається в обов'язковому порядку. Тому процес управління природно розбитий на 5 етапів (кроків). Параметром стану є вік обладнання t. На початку першого року експлуатації обладнання нове і $S_0=0$, у наступні роки $S_{i-1}=t$. Управління на кожному кроці обирається із змінних x^c і $x^з$.

Рівняння станів мають вигляд

$$S_i = \begin{cases} t+1, & якщо\ x_i = x^c \\ 1, & якщо\ x_i = x^з, \quad i = 1,2,3,4 \end{cases},$$

звідки видно, що якщо до початку і-го кроку $S_{i-1}=t$, і експлуатація обладнання триває, тобто $x_i=x^c$, то вік обладнання збільшується на одиницю. Якщо ж обладнання замінюється на нове, то на початку і-го кроку його вік t=0, а наприкінці - t=1.

Визначимо показник ефективності і-го кроку з урахуванням того, що якщо експлуатація обладнання триває, тобто $x_i=x^c$, то мають місце витрати тільки на утримання обладнання c(t)=600(t+1). Якщо ж обладнання замінюється новим, тобто $x_i=x^з$, то витрати включають P=4000 грн. на покупку нового обладнання за винятком ліквідаційної вартості проданого обладнання, вік якого t років $p(t)=P*2^{-t}$ = 4000*2^{-t}, і витрат на утримання нового обладнання протягом року $c(t)$=600, тобто 4000+600-4000*2^{-t}. Маємо

$$w_i = \begin{cases} 600(t+1), & якщо\ x_i = x^c \\ 4000+600-4000*2^{-t}, & якщо\ x_i = x^з, \quad i = 1,2,3,4 \end{cases}.$$

Нехай w^*_i – умовні оптимальні витрати на експлуатацію обладнання з і-го до 5-го кроку, за умови, що до початку і-го кроку обладнання має вік t років. Запишемо для функцій w^*_i рекурентні співвідношення (рівняння Беллмана):

$$w^*_5 = \min \begin{cases} 600(t+1) - 4000*2^{-(t+1)}, & якщо\ x_5 = x^c \\ 4000+600-4000*2^{-t} - 4000*2^{-(t+1)}, & якщо\ x_5 = x^з \end{cases},$$

де 4000*$2^{-(t+1)}$ – ліквідаційна вартість устаткування віком t років.

Або

$$w*_i = \min\begin{cases} 600(t+1) - w*_{i+1}(t+1), & \text{якщо } x_i = x^c \\ 4000 + 600 - 4000*2^{-t} + w*_{i+1}(1), & \text{якщо } x_i = x^з \end{cases},$$

де i=4,3,2,1.

Складемо таблицю, помістивши до неї вихідні розрахункові дані.

Таблиця 4.2 - Розрахункові дані задачі

Вік обладнання, років	**Вартість утримання c(t)=600(t+1), грн.**	**Залишкова вартість p(t)=P*2^{-t}, грн.**
0		4000
1	600	2000
2	1200	1000
3	1800	500
4	2400	250
5	3000	125

Зобразимо схему процесу управління на рисунку 4.1. На осі абсцис відкладатимемо етапи рішення і (5 кроків), а на осі ординат - вік обладнання *t*. У кружках цифри вказують стан обладнання *S*.

На початку першого року є нове обладнання (*t*=0), і протягом першого року воно знаходиться в експлуатації. Наприкінці першого року вік *t*=1, ми можемо або зберегти експлуатацію обладнання (*С*), і тоді наприкінці другого року його вік стане *t*=2, або замінити його новим (*З*), і тоді наприкінці другого року його вік *t*=1. Такий самий підхід використовується на початку кожного наступного року. Якщо однолітнє обладнання замінюється на початку другого або третього років, то обладнання, що його замінило, на початок наступного року так само буде однолітнім. Наприкінці п'ятого року все обладнання продається.

Зробимо розрахунок витрат у відповідності із схемою процесу управління, що наведена на рисунку 4.1, починаючи з останнього 5-го кроку та переміщаючись до початку процесу. У кружках запишемо значення сумарних витрат, а над стрілками - значення витрат управління.

На 5-му етапі ми маємо 5 можливих станів обладнання S_5=(1, 2, 3, 4, 5). Оскільки все обладнання продається, запишемо в кружки відповідну ліквідаційну вартість із від’ємним знаком, оскільки критерієм ефективності є витрати (рис. 4.2).

Визначимо витрати, зроблені при переході із стану S_4 у стан S_5, потім для кожного стану обчислимо сумарні витрати, виберемо з них найменші та запишемо в кружки 4-го етапу. Умовні оптимальні управління виділимо подвійними лініями. Результати розрахунків помістимо в таблицю 4.3.

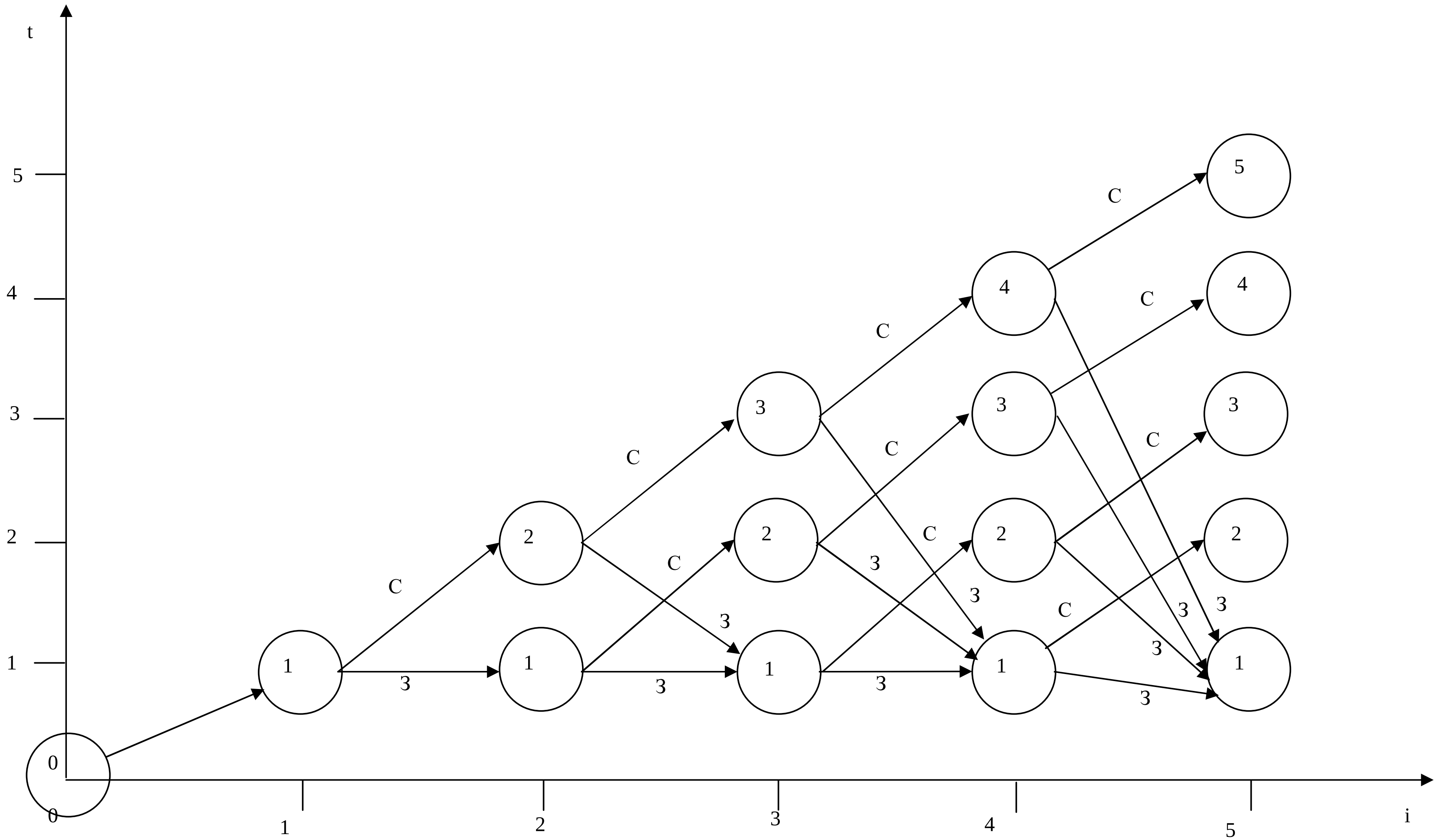


Рисунок 4.1 – Схема процесу управління

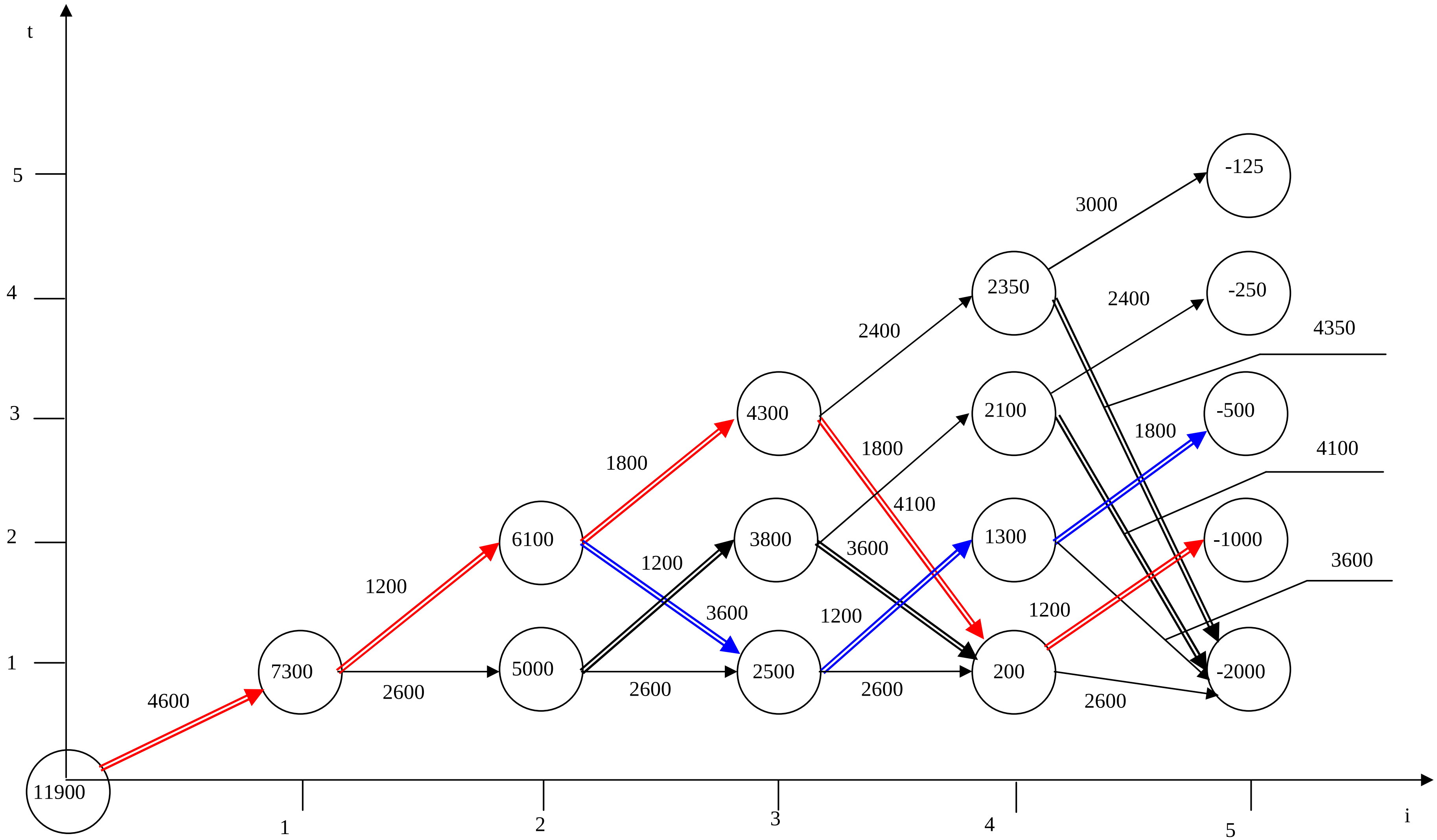


Рисунок 4.2 – Розв’язання задачі

Таблиця 4.3 - Етап 5

Вік t (стан S_i)	**Управління x^c $w_5=600*(t+1)$**	**Управління $x^з$ $w_5=4000+600-$ $-4000*2^{-t}$**	**Сумарні витрати, $w_5(t)$**	**Оптимум**	
				$w^*_5(t)$	**Рі-шен-ня**
1	600(1+1)=1200	4000+600- $-4000*2^{-1}$=2600	min{1200-1000; 2600-2000}=200	200	С
2	600(2+1)=1800	4000+600- $-4000*2^{-2}$=3600	min{1800-500; 3600-2000}=1300	1300	С
3	600(3+1)=2400	4000+600- $-4000*2^{-3}$=4100	min{2400-250; 4100-2000}=2100	2100	З
4	600(4+1)=3000	4000+600- $-4000*2^{-4}$=4350	min{3000-125; 4350-2000}=2350	2350	З

На 4-му етапі ми маємо 4 можливі стани обладнання S_4=(1, 2, 3, 4). Витрати, зроблені при переході із стану S_3 у стан S_4, і умовні оптимальні управління наведені в таблиці 4.4.

Таблиця 4.4 - Етап 4

Вік t (стан S_i)	**Управління x^c $w_4=600*(t+1)$**	**Управління $x^з$ $w_4=4000+600-4000*2^{-t}$**	**Сумарні витрати, $w_4(t)$**	**Оптимум**	
				$w^*_4(t)$	**Рі-шення**
1					
2	600(1+1)=1200	4000+600- $-4000*2^{-1}$=2600	1300+1200=2500 або 200+2600=2800	2500	С
3	600(2+1)=1800	4000+600- $-4000*2^{-2}$=3600	2100+1800=3900 або 3600+200=3800	3800	З
4	600(3+1)=2400	4000+600- $-4000*2^{-3}$=4100	2350+2400=4750 або 4100+200=4300	4300	З

На 3-му етапі ми маємо 3 можливі стани обладнання S_3=(1, 2, 3). Витрати, зроблені при переході із стану S_2 у стан S_3, і оптимальні рішення наведені в таблиці 4.5.

Таблиця 4.5 - Етап 3

Вік t (стан S_i)	**Управління x^c $w_3=600*(t+1)$**	**Управління $x^з$ $w_3=4000+600-$ $-4000*2^{-t}$**	**Сумарні витрати, $w_3(t)$**	**Оптимум**	
				$w^*_3(t)$	**Рі-шення**
1					
2	600(1+1)=1200	4000+600- $-4000*2^{-1}$=2600	3800+1200=5000 або 2500+2600=5100	5000	С
3	600(2+1)=1800	4000+600- $-4000*2^{-2}$=3600	4300+1800=6100 або 2500+3600=6100	6100	С або З

На 2-му етапі - 2 можливих стана обладнання S_2=(1, 2). Витрати, зроблені при переході із стану S_1 у стан S_2, і оптимальні рішення наведені в таблиці 4.6.

Таблиця 4.6 - Етап 2

Вік t (стан S_i)	**Управління x^c w_2=600*(t+1)**	**Управління $x^з$ w_2=4000+600- -4000*2^{-t}**	**Сумарні витрати, $w_2(t)$**	**Оптимум**	
				$w^*_2(t)$	**Рішен-ня**
1					
2	600(1+1)=1200	4000+600- -4000*2^{-1}=2600	6100+1200=7300 або 5000+2600=7600	7300	С

На 1-му етапі є один можливий стан обладнання S_1=1. Витрати, зроблені при переході із стану S_0 у стан S_1, і оптимальне рішення наведені в таблиці 4.7.

Таблиця 4.7 - Етап 1

Вік t (стан S_i)	**Управління x^c w_1=600*(t+1)**	**Управління $x^з$ w_1=4000+600- -4000*2^{-t}**	**Сумарні витрати, $w_1(t)$**	**Оптимум**	
				$w^*_1(t)$	**Рішен-ня**
1	600(0+1)=600		4000+600+7300 = =11900	11900	С

Таким чином, у результаті здійснення умовної оптимізації ми одержали в точці S_0=0 значення мінімальних витрат на експлуатацію обладнання протягом 5 років. Переміщаючись з точки S_0=0 за подвійними стрілками, побудуємо оптимальну траєкторію управління (на рисунку 4.2 відзначена червоним). Вектор оптимального управління має вигляд:

$$x=(x^C, x^C, x^C, x^з, x^C),$$

відповідно до якого обладнання варто експлуатувати протягом перших трьох років, потім замінити новим.

З рисунку 4.2 видно, що задача має альтернативне рівноцінне рішення (відзначене синім):

$$x=(x^C, x^C, x^з, x^C, x^C).$$

Наявність альтернативного оптимуму в даній задачі пов'язана з тим, що в стані S_2=2 цільова функція дорівнює 6100 при обох рішеннях x^C і $x^з$. Таким чином, можна обладнання експлуатувати протягом перших двох років, а потім замінити новим, яке експлуатувати так само два роки.

4.4 Задача про розподіл інвестиційних ресурсів

Суть завдання про розподіл інвестиційних ресурсів збігається до наступного. Припустимо, що на початку кожного з n років необхідно здійснювати інвестування в сумах P_1, P_2,.,P_n. Є два банки, в які можна вкласти капітал. Обидва банки виплачують річний складний відсоток r_1 і r_2 відповідно. Окрім цього, обидва банки виплачують новим інвесторам премії у вигляді відсотка q від вкладеної суми. Відсоток премії може змінюватися і для і-го року складає q_{i1}

для першого банку і q_{i2} - для другого. Премії виплачуються в кінці року і можуть бути інвестовані на початку наступного року в будь-який з банків. Необхідно розробити оптимальну стратегію інвестування на n років.

Відповідно до рекомендацій, розглянутих в п. 4.2, розділимо операцію управління інвестиціями на етапи (кроки). Етап і представимо порядковим номером року і, $i=\overline{1,n}$. Кроковим управлінням на і-му кроці є суми інвестицій I_{i1} і I_{i2} у перший і другий банк відповідно. Стан системи на і-му кроці визначається сумою грошових коштів x_i, які можуть бути інвестовані. Сума грошових коштів x_i включає тільки нові грошові кошти та преміальні відсотки за інвестиції, зроблені впродовж попереднього (і-1)-го року $x_i = I_{i1} + I_{i2}$.

Отже, стан на першому кроці визначається сумою грошових коштів

$$x_1=P_1.$$

На наступних кроках

$$x_i = P_i + q_{i-1,1}I_{i-1} + q_{i-1,2}(x_{i-1} - I_{i-1}) =$$
$$= P_i + (q_{i-1,1} - q_{i-1,2})I_{i-1} + q_{i-1,2}x_{i-1}. \qquad (4.13)$$

Завдання полягає в тому, щоб максимізувати суму накопичених коштів $S=s_1+s_2+...+s_n$, де s_i - сума коштів, накопичених до кінця і-го року:

$$s_i = I_i(1+r_1)^{n+1-i} + (x_i - I_i)(1+r_2)^{n+1-i}, \quad i=\overline{1,i-1}; \qquad (4.14)$$

$$s_n = I_n(1+r_1+q_{n1}) + (x_n - I_n)(1+r_2+q_{n2}). \qquad (4.15)$$

Рекурентне рівняння зворотного прогону для обчислення оптимальної суми інвестицій $f_i(x_i)$ для інтервалу від і-го до n-го року має вигляд

$$f_i(x_i) = \max_{0<Ii<xi}\{s_i + f_{i+1}(x_{i+1})\}, \quad i=\overline{1,i-1}, \qquad (4.16)$$

де $f_{i+1}(x_{i+1})\equiv 0$.

Приклад 4.3. Нехай на початку кожного з 4 років необхідно здійснювати інвестування в сумах P_1=4000 грн. P_2=P_3=P_4=2000 грн. Є два банки, в які можна вкласти капітал. Перший банк виплачує річний складний відсоток, що становить 8%, а другий - на 0,2% нижче за перший. Обидва банки виплачують новим інвесторам премії у вигляді відсотку q від вкладеної суми. Відсоток премії для першого банку дорівнює за роками відповідно 1,8%, 1,7%, 2,1% і 2,5%. Відсоток премії другого банку на 0,5% вищий за першого. Премії виплачуються наприкінці року і можуть бути інвестовані на початку наступного року в будь-який з банків. Необхідно розробити оптимальну стратегію інвестування на чотири роки.

Розв'язання

Визначимо елементи моделі динамічного програмування:

P_1=4000 грн. P_2=P_3=P_4=2000 грн.;
r_1=0,08; r_2=0,078;
q_{11}=0,018, q_{21}=0,017, q_{31}=0,021, q_{41}=0,025;
q_{12}=0,023, q_{22}=0,022, q_{32}=0,026, q_{42}=0,03.

Етап 4.

Відповідно до (4.16) і з урахуванням того, що $f_{i+1}(x_{i+1})\equiv 0$, оптимальна сума інвестицій на 4-му кроці визначиться таким чином:

$$f_4(x_4) = \max_{0<I4<x4}\{s_4\},$$

де сума, накопичена на кінець 4-го року, складатиме

$$s_4 = I_4(1{,}08+0{,}025)+(x_4-I_4)(1{,}078+0{,}03) = I_4 * 1{,}105+(x_4-I_4)*1{,}108 =$$
$$= I_4 * 1{,}105 + x_4 * 1{,}108 - I_4 * 1{,}108 = 1{,}108x_4 - 0{,}003I_4.$$

Оскільки другий доданок з від’ємним знаком, очевидно, що s_4 має максимум при $1{,}108x_4$:

Таблиця 4.8 - Етап 4

Стан	**Оптимальне рішення**	
	$f_4(x_4)$	**I^*_4**
x_4	$1{,}108x_4$	0

Етап 3.

Оптимальна сума інвестицій на 3-му кроці

$$f_3(x_3) = \max_{0<I3<x3}\{s_3 + f_4(x_4)\},$$

де сума, накопичена до кінця 3-го року

$$s_3 = I_3(1{,}08)^{4+1-3} + (x_3 - I_3)(1{,}078)^{4+1-3} = 1{,}08^2 I_3 - 1{,}078^2 I_3 + 1{,}078^2 x_3$$
$$= 1{,}00432I_3 + 1{,}1621x_3,$$

а сума грошових коштів, які можуть бути інвестовані на 4-му етапі, складе

$$x_4=P_4-0{,}005I_3+q_{32}x_3 = 2000-0{,}005I_3+0{,}026x_3.$$

Тоді оптимальна сума інвестицій на 3-му кроці

$$f_3(x_3) = \max_{0<I3<x3}\{s_3 + f_4(x_4)\} = \max_{0<I3<x3}\{1{,}00432I_3 + 1{,}1621x_3 + 1{,}108(2000 - 0{,}005I_3 + 0{,}026x_3)\} =$$
$$= \max_{0<I3<x3}\{1{,}00432I_3 + 1{,}1621x_3 + 2216 - 1{,}00554I_3 + 0{,}0288x_3\} =$$
$$= \max_{0<I3<x3}\{2216 - 0{,}00122I_3 + 1{,}1909x_3\}.$$

Оскільки другий доданок з від’ємним знаком, очевидно, що s_3 має максимум при $2216+1{,}1909x_3$:

Таблиця 4.9 - Етап 3

Стан	**Оптимальне рішення**	
	$f_3(x_3)$	**I^*_3**
x_3	$2216+1{,}1909x_3$	0

Етап 2.

Оптимальна сума інвестицій на 2-му кроці

$$f_2(x_2) = \max_{0<I2<x2}\{s_2 + f_3(x_3)\},$$

де сума, накопичена до кінця 2-го року

$$s_2 = I_2(1{,}08)^{4+1-2} + (x_2 - I_2)(1{,}078)^{4+1-2} = I_2(1{,}08)^3 + (x_2 - I_2)(1{,}078)^3 =$$
$$= 1{,}2597I_2 - 1{,}2527I_2 + 1{,}2527x_2 = 0{,}006985I_2 + 1{,}2527x_2,$$

а сума грошових коштів, які можуть бути інвестовані на 3-му етапі, складе

$$x_3=P_3-0{,}005I_2+q_{22}x_2 = 2000-0{,}005I_2+0{,}022x_2.$$

Тоді оптимальна сума інвестицій на 2 кроці

$$f_2(x_2) = \max_{0<I2<x2}\{s_2 + f_3(x_3)\} =$$

$$= \max_{0<I2<x2}\{0{,}006985\mathrm{I}_2 + 1{,}2527x_2 + 2216 + 1{,}1909(2000 - 0{,}005\mathrm{I}_2 + 0{,}022x_2)\} =$$

$$= \max_{0<I2<x2}\{0{,}006985\mathrm{I}_2 + 1{,}2527x_2 + 2216 + 2381{,}8 - 0{,}005955I_2 + 0{,}0262x_2\} = \cdot$$

$$= \max_{0<I2<x2}\{0{,}001031I_2 + 1{,}2789x_2 + 4597{,}8\}$$

Покладемо $I_2=x_2$, тоді s_2 має максимум при $4597{,}8+1{,}27996x_2$:

Таблиця 4.10 - Етап 2

Стан	**Оптимальне рішення**	
	$f_2(x_2)$	I^*_2
x_2	$4597{,}8+1{,}27996x_2$	x_2

Етап 1.

Оптимальна сума інвестицій на 1-му кроці

$$f_1(x_1) = \max_{0<I1<x1}\{s_1 + f_2(x_2)\},$$

де сума, накопичена до кінця 1-го року

$$\mathrm{s}_1 = \mathrm{I}_1(1{,}08)^{4+1-1} + (\mathrm{x}_1 - \mathrm{I}_1)(1{,}078)^{4+1-1} = 1{,}3605\mathrm{I}_1 - 1{,}35044\mathrm{I}_1 + 1{,}35044\mathrm{x}_1 =$$

$$= 0{,}01005\mathrm{I}_1 + 1{,}35044\mathrm{x}_1,$$

а сума грошових коштів, які можуть бути інвестовані на 2-му етапі, складе

$$x_2=P_2-0{,}005I_1+q_{12}x_1 = 2000-0{,}005I_1+0{,}023x_1.$$

Тоді оптимальна сума інвестицій на 1 кроці складе

$$f_1(\mathrm{x}_1) = \max_{0<I1<x1}\{s_1 + f_2(x_2)\} =$$

$$= \max_{0<I1<x1}\{0{,}01005I_1 + 1{,}35044x_1 + 4597{,}8 + 1{,}27996(2000 - 0{,}005I_1 + 0{,}023x_1)\} =$$

$$= \max_{0<I1<x1}\{0{,}01005I_1 + 1{,}35044x_1 + 4597{,}8 + 2559{,}92 - 0{,}0064I_1 + 0{,}02944x_1)\} = \cdot$$

$$= \max_{0<I2<x2}\{0{,}00365I_1 + 1{,}3799x_1 + 7157{,}72\}.$$

Вважаємо $I_1=x_1$, тоді s_1 має максимум при $7157{,}72+1{,}3835x_1$:

Таблиця 4.11 - Етап 1

Стан	**Оптимальне рішення**	
	$\mathbf{f_1(x_1)}$	$\mathbf{I^*_1}$
$x_1=4000$	$7157{,}72+1{,}3835x_1$	4000

Проведемо обчислення у зворотному напрямі:
x_2= 2000-0,005*4000+0,023*4000=2072 грн.,
x_3= 2000-0,005*2072+0,022*2072=2035,2 грн.,
x_4= 2000-0,005*0+0,026*2035,2=2052,9 грн.
Складемо оптимальну стратегію, наведену у таблиці 4.12.

Таблиця 4.12 - Оптимальна стратегія інвестування

Рік	**Оптимальне рішення**	**Рішення інвестора**	**Накопичення**
1	$I^*_1=x_1$	Вкласти x_1=4000 грн. в перший банк	s_1=0,01005*4000+1,35044*4000= =5442 грн.
2	$I^*_2=x_2$	Вкласти x_2=2072 грн. в перший банк	s_2=0,006985*2072+1,2527*2072= =2610,07 грн.
3	$I^*_3=0$	Вкласти x_3=2035,2 грн. в другий банк	s_3=0,00432*0+1,1621*2035,2= =2365,11 грн.
4	$I^*_4=0$	Вкласти x_4=2052,9 грн. в другий банк	s_4=0,003*0+1,108*2052,9=2274,61 грн.
	Всього:		12691,7 грн.

Контрольні запитання

1. Поясніть сутність завдань про заміну обладнання. За якими ознаками їх класифікують?

2. За якими способами визначають період експлуатації обладнання, після якого має бути зроблена його заміна?

3. Охарактеризуйте особливості методу динамічного програмування. Для розв’язання яких задач він призначений

4. Поясніть, що являють собою аддитивний та мультиплікативний критерії.

5. Сформулюйте принцип оптимальності Беллмана та поясніть відсутність післядії.

6. Поясніть що називають умовним кроковим управлінням та умовним оптимальним управлінням.

7. Якими елементами завдання про заміну зумовлені стани системи. Запишіть рівняння станів.

8. За якими критеріями виконують оптимізацію задачі про заміну обладнання?

9. Запишіть рекурентні співвідношення, використовувані при розв’язанні задачі про заміну обладнання та поясніть їх.

ЗМ 2 Стохастичні моделі і методи

ТЕМА 5
ЗАДАЧІ МАСОВОГО ОБСЛУГОВУВАННЯ

5.1 Сукупність задач масового обслуговування

Під час дослідження складних систем зустрічаються системи, що призначені для багаторазового використання при вирішенні однотипних задач. Процеси, що виникають при цьому, називають процесами обслуговування, а системи - системами масового обслуговування (СМО). Системи масового обслуговування - це такі системи, до яких у випадкові моменти часу надходять **заявки на обслуговування**, при цьому заявки, що надійшли, обслуговуються наявними у розпорядженні системи **каналами** обслуговування. Обслуговування заявки відбувається протягом певного часу $t_{об}$, величина якого випадкова. Потім канал звільняється і готовий до прийому наступної заявки. Випадковий характер потоку заявок і часу обслуговування $t_{об}$ призводить до того, що в певні періоди часу на вході СМО нагромаджується надто велика кількість заявок. Вони або стають у чергу, або залишають СМО без обслуговування. В інші ж періоди часу СМО працюватиме з недовантаженням або простоюватиме. Стан СМО змінюється стрибком у моменти появи нової заявки або закінчення обслуговування, або в момент, коли заявка залишає чергу.

Прикладами систем масового обслуговування можуть бути станції технічного обслуговування автомобілів; персональні комп'ютери, що виконують вирішення тих або інших задач; аудиторські фірми; відділи податкових інспекцій, які займаються прийманням і перевіркою поточної звітності підприємств; квиткові каси; магазини; перукарні; телефонні станції та ін.

Предметом теорії масового обслуговування є побудова математичних моделей, що зв'язують задані умови роботи СМО (число каналів, їхню продуктивність, характер потоку заявок та ін.) з показниками ефективності СМО, що описують її здатність справлятися з потоком заявок. Показниками ефективності СМО є: середнє число заявок, що обслуговуються в одиницю часу; середнє число заявок у черзі; середній час очікування обслуговування; середнє число зайнятих каналів; імовірність відмови в обслуговуванні без очікування; імовірність того, що число заявок в черзі перевищить певне значення та ін.

5.2 Основні поняття та визначення задач масового обслуговування

Основними компонентами системи масового обслуговування будь-якого виду є вхідний потік вимог або заявок на обслуговування; дисципліна черги; механізм обслуговування.

Вхідний потік вимог. Для опису вхідного потоку потрібно задати імовірнісний закон, що визначає послідовність моментів надходження вимог на обслуговування і вказати кількість таких вимог в кожному черговому надходженні. При цьому, як правило, оперують поняттям «імовірнісний розподіл моментів надходження вимог». Тут можуть надходити як одиничні, так і групові вимоги (надходять групами). В останньому випадку зазвичай йдеться про систему обслуговування з паралельно-груповим обслуговуванням.

Дисципліна черги - це важливий компонент системи масового обслуговування, він визначає принцип, відповідно до якого заявки, що надходять на вхід обслуговуючої системи, підключаються з черги до процедури обслуговування. Найчастіше використовують дисципліни черги, зумовлені наступними правилами:

- перший прийшов - перший обслуговується;
- прийшов останнім - обслуговується першим;
- випадковий відбір заявок;
- відбір заявок за критерієм пріоритетності;
- обмеження часу очікування моменту настання обслуговування (має місце черга з обмеженим часом очікування обслуговування, що асоціюється з поняттям «припустима довжина черги»).

Механізм обслуговування визначається характеристиками самої процедури обслуговування і структурою обслуговуючої системи. До характеристик процедури обслуговування належать: тривалість обслуговування і кількість вимог, що задовольняються в результаті виконання кожної такої процедури.

Для аналітичного опису характеристик обслуговування оперують поняттям «імовірнісний розподіл часу обслуговування». Слід зазначити, що час обслуговування заявки залежить від характеру самої заявки або вимог клієнта і від стану та можливостей обслуговуючої системи.

Структура СМО визначається кількістю і взаємним розташуванням каналів обслуговування (механізмів, приладів та ін.). Система може мати не один канал обслуговування, а кілька. Така система здатна обслуговувати одночасно кілька вимог. У цьому разі всі канали обслуговування пропонують ті самі послуги і, отже, можна стверджувати, що має місце паралельне обслуговування. Система може складатися з кількох різнотипних каналів обслуговування, через які повинна пройти кожна обслуговувана заявка, тобто в обслуговуючій системі процедури обслуговування заявок реалізуються послідовно. Механізм обслуговування визначає характеристики вихідного (обслуженого) потоку заявок.

5.3 Характеристика найпростішого потоку вимог

СМО ділять на два основних класи: СМО з відмовами і СМО з очікуванням (з чергою). У СМО з відмовами заявка, що надійшла в момент, коли всі канали зайняті, отримує відмову, залишає СМО і надалі не бере участі в процесі обслуговування (наприклад, заявка на телефонну розмову в момент, коли всі канали зайняті, отримує відмову і залишає СМО не обслуженою). У СМО з очікуванням заявка, що прийшла в момент, коли всі канали зайняті, не йде, а стає в чергу на обслуговування.

СМО з очікуванням підрозділяються на різні види залежно від того, як організована черга: з обмеженою або необмеженою довжиною черги, з обмеженим часом очікування та ін.

Математичний аналіз СМО істотно полегшується, якщо процес її роботи - марківський. Випадковий характер потоку заявок, а також, у загальному випадку, і тривалості обслуговування $t_{об}$ призводить до того, що в системі масового обслуговування протікає випадковий процес.

Випадковий процес, що відбувається в системі, називається марківським, якщо для кожного моменту часу t_0 імовірність будь-якого стану системи в майбутньому залежить тільки від її стану в даний момент часу t_0 і не залежить від того, яким способом система прийшла в цей стан.

На практиці марківські процеси в чистому вигляді не зустрічаються, але нерідко доводиться мати справу з процесами, для яких впливом передісторії можна зневажити, і для їх вивчення можна застосовувати марківські моделі.

Велике значення мають марківські випадкові процеси з дискретними станами і безперервним часом. Процес називається процесом з **дискретними станами**, якщо його можливі стани S_1, S_2, ...S_n можна заздалегідь перелічити, а перехід системи зі стану до стану відбувається миттєво (стрибком).

Процес називається процесом з **безперервним часом**, якщо моменти можливих переходів системи зі стану до стану не фіксовані заздалегідь, а випадкові.

Процес роботи СМО являє собою випадковий процес з дискретними станами і безперервним часом. Це означає, що стан СМО змінюється стрибком у випадкові моменти появи якихось подій (наприклад, приходу нової заявки, закінчення обслуговування та ін.).

У випадку немарківських процесів задачі дослідження систем масового обслуговування значно ускладнюються і вимагають застосування статистичного моделювання і числових методів з використанням ЕОМ.

При аналізі випадкових процесів з дискретними станами зручно користуватися спеціальною геометричною схемою - графом станів. Зазвичай стани системи зображують прямокутниками (або кружками), а можливі переходи з одного стану в інший - стрілками (орієнтованими дугами), що з'єднують стани.

Для математичного опису марківського випадкового процесу з дискретними станами і безперервним часом, що протікає в СМО, познайомимося з одним з важливих понять - поняттям потоку подій.

Потік подій – це послідовність однорідних подій, що йдуть одна за одною у випадкові моменти часу (наприклад, потік викликів на телефонній станції, потік відмов ЕОМ, потік покупців та ін.). Потік характеризується інтенсивністю λ - частотою появи подій або середнім числом подій, що надходять у СМО в одиницю часу.

Потік подій називається **регулярним**, якщо події йдуть одна за одною через рівні проміжки часу.

Потік подій називається **стаціонарним**, якщо його імовірнісні характеристики не залежать від часу. Зокрема, інтенсивність стаціонарного потоку є величиною постійною: $\lambda(t)=\lambda=\text{const}$.

Потік подій називається **потоком без післядії**, якщо для будь-яких двох ділянок часу τ_1 і τ_2, що не перетинаються, число подій, що потрапляють на одну з них, не залежить від числа подій, що потрапляють на інші.

Потік подій називається **ординарним**, якщо імовірність попадання на малу (елементарну) ділянку часу Δt двох і більше подій зневажливо мала порівняно з імовірністю попадання однієї події.

Нагадаємо, що потік подій називається найпростішим, якщо він одночасно стаціонарний, ординарний і не має післядії. Назва "найпростіший" пояснюється тим, що СМО з найпростішими потоками має найбільш простий математичний опис. Для найпростішого потоку число подій m, що потрапляють на довільну ділянку часу τ, розподілене за законом Пуассона

$$P_{\mathrm{m}}(\tau)=\frac{(\lambda\tau)^m}{m!}e^{-\lambda\tau},$$

для якого математичне сподівання випадкової величини m дорівнює її дисперсії:

$$a = \sigma^2 = \lambda\tau.$$

Зокрема, імовірність того, що за час τ не відбудеться жодної події (m=0), дорівнює

$$P_0 = \mathrm{e}^{-\lambda\tau}.$$

Інтервал часу між довільними двома сусідніми подіями найпростішого потоку T розподілений за експонентним законом:

$$F(t) = P(T < \mathrm{t}) = 1 - \mathrm{e}^{-\lambda \mathrm{t}}.$$

Щільність імовірності випадкової величини T є похідною її функції розподілу, тобто

$$f(t)=F'(t)= \lambda \mathrm{e}^{-\lambda \mathrm{t}}.$$

Математичне сподівання є величиною, зворотною λ, і дорівнює середньому квадратичному відхиленню випадкової величини T:

$$m = \sigma = \frac{1}{\lambda}.$$

Найважливіша властивість експонентного розподілу (притаманна тільки експонентному розподілу) полягає в наступному:

> **якщо проміжок часу, розподілений за експонентним законом, уже тривав якийсь час τ, то це ніяк не впливає на закон розподілу частини, що залишилася, проміжку (Т-τ): він буде таким самим, як і закон розподілу всього проміжку Т.**

Інакше кажучи, для інтервалу часу T між двома послідовними сусідніми подіями потоку, що має експонентний розподіл, будь-які відомості про те, скільки часу тривав цей інтервал, не впливають на закон розподілу частини, що залишилася. Ця властивість експонентного закону являє собою, по суті, інше формулювання для «відсутності післядії» - основної властивості найпростішого потоку.

Як «міру випадковості» невід’ємної випадкової величини використовують коефіцієнт варіації:

$$V = \frac{\sigma}{m_t}.$$

Для **експонентного** розподілу, а отже для найпростішого потоку, коефіцієнт варіації дорівнює одиниці. Для регулярного потоку, в якого інтервал між подіями невипадковий і σ=0, коефіцієнт варіації дорівнює нулю. Для більшості реальних потоків подій коефіцієнт варіації інтервалів між подіями лежить у межах

$$0 \le V \le 1.$$

Він може служити мірою ступеня регулярності потоку. Чим ближче V до нуля, тим регулярніше потік. Найпростіший потік найменш регулярний з усіх потоків подій.

5.4 Рівняння для імовірностей станів СМО

Як приклад розглянемо одноканальну СМО. Очевидно, що така СМО може знаходитись в одному з двох станів:

S_0 – канал вільний;

S_1 – канал зайнятий.

Якщо всі потоки подій, що переводять систему з одного стану в інший, найпростіші, то процес, що відбувається в системі, буде марківським. У цьому випадку його можна описати звичайними диференціальними рівняннями.

Побудуємо граф станів для одноканальної СМО (рис. 5.1).

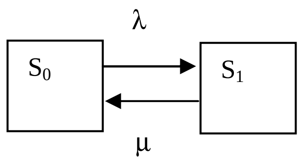


Рисунок 5.1 - Спрямований граф одноканальної СМО з відмовами

Зі стану S_0 у стан S_1 систему переводить потік заявок, інтенсивність якого λ, а зі стану S_1 у стан S_0 систему переводить потік обслуговувань, інтенсивність якого μ. Нехай імовірності станів системи становитимуть p_0 і p_1 відповідно. Очевидно, що для будь-якого моменту часу сума імовірностей станів є імовірністю достовірної події

$$p_0(t) + p_1(t) = 1. \tag{5.1}$$

Якщо є граф станів, можна визначити імовірності станів системи $p_0(t)$ і $p_1(t)$ як функції часу. Для цього складають і вирішують рівняння Колмогорова - диференціальні рівняння, в яких невідомими функціями є імовірності станів.

Оскільки потоки заявок і обслуговувань – найпростіші, інтенсивність λ = const, і час між заявками розподілений за експонентним законом

$$f(t) = \lambda * e^{-\lambda t}.$$

Для процесу обслуговувань також μ= const, тривалість обслуговування $t_{об}$ розподілена за експонентним законом, тому

$$f(t_{об}) = \mu * e^{-\mu t}.$$

Розглянемо одну з імовірностей станів, наприклад, $p_0(t)$. Це імовірність того, що в момент часу t система знаходиться в стані S_0. Знайдемо імовірність $p_0(t+\Delta t)$, тобто імовірність того, що в момент часу $(t+\Delta t)$ канал вільний. У момент часу $(t+\Delta t)$ система знаходиться в стані S_0 в двох випадках:

а) якщо в момент часу t система знаходилась у стані S_0 і за час Δt вона не вийшла з цього стану;

б) якщо система в момент часу t була в стані S_1 і за час Δt перейшла в стан S_0.

Знайдемо імовірність варіанту а). Імовірність, що система в момент часу t знаходилась у стані S_0 дорівнює $p_0(t)$. Імовірність того, що за час Δt не надійшло жодної заявки визначиться за законом Пуассона

$$\frac{(\lambda\Delta t)^0}{0!}e^{-\lambda\Delta t}=e^{-\lambda\Delta t}, \tag{5.2}$$

розкладання в ряд Тейлора дає приблизно

$$e^{-\lambda\Delta t} = 1-\lambda\Delta t+\lambda^2\Delta t^{2/2} - \ldots \approx 1-\lambda\Delta t.$$

Таким чином, імовірність дотримання умови "а)" за теоремою множення дорівнює:

$$p_0(t)*(1-\lambda\Delta t). \tag{5.3}$$

Знайдемо імовірність варіанту б). Імовірність, що система в момент часу t знаходилась у стані S_1 дорівнює $p_1(t)$. Імовірність того, що за час Δt канал звільнився, визначиться за законом Пуассона через імовірність протилежної події

$$1-\frac{(\mu\Delta t)^0}{0!}e^{-\mu\Delta t}=1-e^{-\mu\Delta t}\approx\mu\Delta t \tag{5.4}$$

Тоді імовірність дотримання умови "б)" за теоремою множення

$$p_1(t)*\mu\Delta t. \tag{5.5}$$

Таким чином, імовірність того, що в момент часу $(t+\Delta t)$ система знаходиться в стані S_0 (канал вільний), визначиться за теоремою додавання імовірностей

$$p_0(t+\Delta t) = p_0(t) * (1-\lambda\Delta t) + p_1(t) * \mu\Delta t. \tag{5.6}$$

Перенесемо $p_0(t)$ у ліву частину, розділимо на Δt і в границі при $\Delta t\to 0$ отримаємо

$$\frac{dp_0(t)}{dt}=-p_0(t)\lambda+p_1(t)\mu. \tag{5.7}$$

Диференціальне рівняння для стану S_1 дістанемо, міркуючи аналогічно:

$$\frac{dp_1(t)}{dt}=p_0(t)\lambda-p_1(t)\mu. \tag{5.8}$$

Спрямований граф, наведений на рисунку 5.1, ілюструє процес зміни станів. Його вершини відповідають станам, а ребра - можливим переходам з одного стану в інший.

Якщо є спрямований граф станів, то систему диференціальних рівнянь для імовірностей станів p_k (k=0,1,2,…) можна записати, користуючись наступним правилом:

- у лівій частині кожного диференціального рівняння знаходиться похідна імовірності k-го стану, а в правій - стільки доданків, скільки ребер зв'язано безпосередньо з даним станом;

- якщо ребро закінчується в даному стані, то доданок має знак плюс, якщо починається з даного стану, - мінус;

- кожен доданок дорівнює добутку інтенсивності потоку подій, що переводить елемент або систему за даним ребром в інший стан, на імовірність того стану, з якого починається ребро.

Щоб вирішити рівняння Колмогорова та знайти імовірності станів, необхідно насамперед задатися початковими умовами. Наприклад, у момент часу $t = 0$ система знаходилась в стані S_1. Тоді початкові умови $p_0(0) = 1$, $p_1(0) = 0$. Подібні рівняння можна вирішувати аналітично або чисельно - вручну або на ЕОМ. У результаті вирішення визначаються імовірності станів як функції часу:

$$p_0(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t};$$

$$p_1(t) = \frac{\lambda}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}. \qquad (5.9)$$

При $t \to \infty$ в системі встановлюється граничний стаціонарний режим, в ході якого система випадковим чином змінює свої стани, але їхні імовірності вже не залежать від часу. Ці імовірності називаються **фінальними імовірностями станів**. Фінальну імовірність стану розуміють як середній відносний час знаходження системи в цьому стані. Наприклад, якщо система має два стани, фінальні імовірності яких 0,3 і 0,7, то це означає, що система в середньому 30% часу проводить у стані S_0.і 70% - у стані S_1. Можна оцінити середню ефективність роботи системи. Нехай в стані S_0 вона дає нульовий дохід, а в стані S_1 дохід становить 15 грн. Тоді

E = 0,3*0+0,7*15=10,5 грн.

Для обчислення фінальних імовірностей в системі диференціальних рівнянь необхідно покласти похідні імовірностей рівними нулю і розв'язати отриману систему алгебраїчних рівнянь. Можна також з графа станів відразу записати систему алгебраїчних рівнянь для фінальних імовірностей станів. Для цього користуються правилом:

- ліворуч стоїть фінальна імовірність стану p_k, помножена на сумарну інтенсивність всіх потоків, що переводять систему з даного стану в інші;

- праворуч стоїть сума добутків інтенсивностей всіх потоків, що переводять систему в даний стан на імовірності тих станів, з яких ці потоки виходять.

Визначимо характеристики одноканальної системи масового обслуговування з відмовами. Для такої системи імовірність p_0 є відносною пропускною здатністю системи q. Дійсно, p_0 - імовірність того, що в момент t канал вільний і заявка, що надійшла, буде обслуженою. З формули (5.9) при $t \to \infty$ дістанемо:

$$q = p_0 = \frac{\mu}{\lambda+\mu}. \qquad (5.10)$$

Знаючи відносну пропускну здатність q, легко знайти абсолютну. Абсолютна пропускна здатність А - це середнє число заявок, які може обслужити система масового обслуговування в одиницю часу:

$$A = \lambda q = \frac{\lambda\mu}{\lambda+\mu}. \qquad (5.11)$$

Імовірність відмови в обслуговуванні заявки дорівнюватиме імовірності стану «канал зайнятий»:

$$P_{відм} = p_1 = 1 - p_0 = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}. \qquad (5.12)$$

Дану величину $P_{\text{відм}}$ можна інтерпретувати як середню частку необслугованих заявок серед тих, що надійшли.

Приклад 5.1. Одноканальна СМО з відмовами являє собою перукарню, де працює один майстер. Заявка - відвідувач, що прибув в момент, коли майстер зайнятий, - одержує відмову в обслуговуванні. Інтенсивність потоку відвідувачів перукарні λ = 1,0 відвідувач/год. Середня тривалість обслуговування становить 1,8 години. Потік відвідувачів і потік обслуговувань є найпростішими. Потрібно визначити відносну пропускну здатність q; абсолютну пропускну здатність А; імовірність відмови $P_{\text{відм}}$, а також порівняти фактичну пропускну здатність СМО з номінальною, яка б мала місце, якби кожен відвідувач обслуговувався точно 1,8 години і відвідувачі надходили один за одним без перерви.

Розв’язання

Визначимо інтенсивність потоку обслуговувань:

$$\mu = \frac{1}{t_{об}} = \frac{1}{1{,}8} = 0{,}555.$$

Обчислимо відносну пропускну здатність:

$$q = \frac{\mu}{\mu+\lambda} = \frac{0{,}555}{1+0{,}555} = 0{,}356.$$

Величина q=0,356 означає, що система в середньому обслуговуватиме 35,6% відвідувачів, які надходять у перукарню.

Абсолютну пропускну здатність знайдемо за формулою

$$A = q\lambda = 0{,}356 * 1 = 0{,}356.$$

Це означає, що система здатна обслужити в середньому 0,356 відвідувач/год.

Імовірність відмови:

$$P_{відм} = 1 - q = 1 - 0{,}356 = 0{,}644.$$

Це означає, що 64,4% прибулих відвідувачів одержать відмову в обслуговуванні.

Визначимо номінальну пропускну здатність системи:

$$A = \frac{1}{t_{об}} = \frac{1}{1{,}8} = 0{,}555$$ відвідувач/год.

Виявляється, що $А_{\text{ном}}$ в 1,5 рази перевищує фактичну пропускну здатність, обчислену з урахуванням випадкового характеру потоку заявок і часу обслуговування ($\frac{A_{ном}}{A} = \frac{0{,}555}{0{,}356} = 1{,}56$).

5.5 Багатоканальна СМО з відмовами

Розглянемо функціонування системи з відмовами (класична задача Ерланга). Нехай є n-канальна система, яку можна уявити як фізичну систему з кінцевою множиною станів: S_0 – всі канали вільні, S_1 – зайнятий рівно один канал, S_k – зайнято рівно k каналів,…, S_n–зайняті всі n каналів. Граф станів системи показаний на рисунку 5.2:

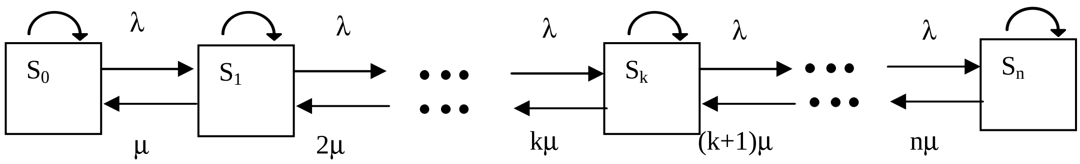


Рисунок 5.2 - Граф станів СМО з відмовами

У верхніх стрілок графа інтенсивності λ, тому що той самий потік заявок переводить систему з одного стану в інший. У нижніх стрілок – інтенсивність обслуговувань μ. У стані S_1 зайнятий один канал, і в стан S_0 система перейде, коли він звільниться. У стані S_2 зайняті два канали, і в стан S_1 система перейде, коли звільняться перший або другий канали, інтенсивність переходу дорівнює $\mu+\mu=2\mu$, і так далі.

Імовірності станів системи $p_k(t)$ для будь-якого моменту часу t можна визначити в такий спосіб. Нехай потік заявок на обслуговування і потік звільнень є найпростішими, інтервал між заявками розподілений за експонентним законом з параметром λ.

Складемо рівняння для $p_k(t)$ $(1\le k\le n)$. Зафіксуємо момент часу t і знайдемо імовірність $p_k(t+\Delta t)$ того, що в момент часу $t+\Delta t$ система буде знаходитись в стані S_k. Ця імовірність обчислюється як імовірність суми трьох подій A, B і C (за числом стрілок, спрямованих в стан S_k на рисунку 5.2): подія A={в момент t система знаходилась в стані S_k і за час Δt заявка не надійшла і жоден канал не звільнився}; подія B={в момент t система знаходилась у стані S_{k-1}, і за час Δt надійшла одна заявка}; подія C={в момент t система знаходилась у стані S_{k+1}, і за час Δt канал звільнився}. За теоремою додавання імовірностей маємо:

$$p_k(t+\Delta t) = P(A) + P(B) + P(C). \tag{5.13}$$

Імовірність події A, тобто того, що за час Δt не надійшла жодна заявка і жоден канал не звільнився, знайдемо відповідно до теореми множення імовірностей:

$$e^{-\lambda\Delta t}\left(e^{-\mu\Delta t}\right)^k = e^{-(\lambda+\mu k)\Delta t},$$

або, зневажуючи величинами малих порядків, маємо:

$$e^{-(\lambda+\mu k)\Delta t} \approx 1-(\lambda+\mu k)\Delta t,$$

тоді

$$P(A) = p_k(t)[1-(\lambda+\mu k)\Delta t]. \tag{5.14}$$

Аналогічно визначимо імовірності подій B і C:

$$P(B) = p_{k-1}(t)\lambda\Delta t\,; \tag{5.15}$$

$$P(C) = p_{k+1}(t)(k+1)\mu\Delta t\,. \tag{5.16}$$

Підставимо (5.14),(5.15) і (5.16) в (5.13) і дістанемо

$$p_k(t+\Delta t) = p_k(t)[1-(\lambda+\mu k)\Delta t] + p_{k-1}(t)\lambda\Delta t + p_{k+1}(t)(k+1)\mu\Delta t\,.$$

Перенесемо $p_k(t)$ в ліву частину рівняння, розділивши на Δt і перейшовши до границі при $\Delta t \to 0$, отримаємо диференціальне рівняння для $p_k(t)$:

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t)\,. \tag{5.17}$$

Аналогічно можна отримати диференціальні рівняння для імовірностей станів системи при k=0 і k=n:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t), \\ \dfrac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t),. \\ \dfrac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - n\mu p_n(t). \end{cases} \tag{5.18}$$

Інтегрування системи рівнянь (5.18) за початкових умов $p_0(0)$ =1, $p_1(0) = \ldots = p_n(0)$ =0 (в початковий момент часу всі канали вільні) дозволяє знайти імовірність кожного з $k = \overline{0,n}$ станів системи.

Імовірність $p_n(t)$ – це імовірність того, що заявка, яка надійшла в момент часу t, застане всі канали зайнятими, тобто одержить відмову. Імовірність $q(t)=1-p_n(t)$ є відносною пропускною здатністю системи. Для даного t це є відношення середнього числа обслужених за одиницю часу заявок до середнього числа поданих.

Замінивши в рівняннях Колмогорова (5.18) похідні імовірностей станів нульовими значеннями, отримаємо систему лінійних алгебраїчних рівнянь, що описують стаціонарний режим:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0, \\ \ldots \\ \lambda p_{k-1} - (\lambda + k\mu)p_k + (k+1)\mu p_{k+1} = 0, \\ \ldots \\ \lambda p_{n-1} - n\mu p_n = 0. \end{cases} \tag{5.19}$$

Вирішивши систему (5.19) щодо імовірностей, дістанемо для будь-якого k:

$$p_k = \frac{\lambda^k}{k!\mu^k}p_0\,. \tag{5.20}$$

Позначимо відношення $\lambda/\mu=\alpha$ і назвемо його приведеною інтенсивністю потоку заявок, яка є числом заявок, що надходять за середній час обслуговування однієї заявки. При цьому формула (5.20) набуде вигляду

$$p_k = \frac{\alpha^k}{k!}p_0\,. \tag{5.21}$$

Звідки, врахувавши, що $\sum_{k=0}^{n} p_k = 1$, дістанемо:

$$\sum_{k=0}^{n} p_k = p_0 \sum_{k=0}^{n} \frac{\alpha^k}{k!} = 1. \tag{5.22}$$

Виразимо звідси p_0 і підставимо в (5.21)

$$p_k = \frac{\frac{\alpha^k}{k!}}{\sum_{k=0}^{n} \frac{\alpha^k}{k!}}, \tag{5.23}$$

де $k = \overline{0,n}$.

Формула (5.23) називається формулою Ерланга. Вона дає граничний закон розподілу числа зайнятих каналів залежно від характеристик потоку заявок і продуктивності системи обслуговування.

Зазначимо, що незважаючи на те, що формули Ерланга в точності справедливі тільки для найпростішого потоку заявок, ними з відомим наближенням можна користуватися і для потоків з обмеженою післядією, а також для систем з очікуванням, коли термін очікування заявки в черзі малий порівняно з середнім часом обслуговування.

Визначимо характеристики функціонування багатоканальної СМО з відмовами у стаціонарному режимі:

- імовірність відмови:

$$P_{відм} = p_n = \frac{(\alpha)^n}{n!} * p_0, \tag{5.24}$$

тому що заявка одержує відмову, якщо надходить в момент, коли всі n каналів зайняті. Величина $P_{відм}$ характеризує повноту обслуговування вхідного потоку;

- імовірність того, що заявка буде прийнята до обслуговування (вона одночасно є відносною пропускною здатністю системи q, доповнює $P_{відм}$ до одиниці):

$$q = 1 - P_{відм} = 1 - \frac{(\alpha)^n}{n!} * p_0, \tag{5.25}$$

- абсолютна пропускна здатність

$$A = \lambda q = \lambda(1 - P_{відм}), \tag{5.26}$$

- середнє число каналів, зайнятих обслуговуванням ($\bar{k}$), наступне:

$$\bar{k} = \sum_{k=1}^{n} k p_k = \alpha(1 - P_{відм}). \tag{5.27}$$

Величина $\bar{k}$ характеризує ступінь завантаження СМО.

Приклад 5.2. Нехай *n*-канальна СМО являє собою обчислювальний центр з трьома (n = 3) взаємозамінними ПЕОМ для вирішення задач. Потік задач, що надходять на ОЦ, має інтенсивність λ = 1 задача на годину. Середня тривалість обслуговування $\bar{t}_{об}$ = 1,8 години. Потік заявок на вирішення задач і потік обслуговування цих заявок є найпростішими. Потрібно обчислити фінальні імовірності станів ОЦ; імовірності відмови в обслуговуванні заявки; відносну пропускну здатність ОЦ; абсолютну пропускну здатність ОЦ; середнє число зайнятих ПЕОМ, а також визначити, скільки додатково потрібно придбати ПЕОМ, щоб скоротити імовірність відмови в обслуговуванні заявки в 2 рази.

Розв'язання

Визначимо параметр потоку обслуговувань μ:

$$\mu = \frac{1}{\bar{t}_{об}} = \frac{1}{1,8} = 0,555.$$

Приведена інтенсивність потоку заявок $\alpha = \frac{1}{0,555} = 1,8$.

Фінальні імовірності станів знайдемо за формулами Ерланга (5.21)-(5.23):

$$p_1 = \frac{\alpha}{1!} * p_0 = \frac{1,8}{1!} = 1,8 * p_0,$$

$$p_2 = \frac{(\alpha)^2}{2!} * p_0 = 1,62 * p_0,$$

$$p_3 = \frac{(\alpha)^3}{3!} * p_0 = 0,97 * p_0,$$

$$p_0 = \frac{1}{\sum_{k=0}^{3} \frac{(\alpha)^k}{k!}} = \frac{1}{1 + 1,8 + 1,62 + 0,97} \approx 0,186,$$

$$p_1 = 1,8 * 0,186 = 0,344,$$

$$p_2 = 1,62 * 0,186 = 0,301,$$

$$p_3 = 0,97 * 0,186 = 0,18.$$

Імовірність відмови в обслуговуванні заявки

$$P_{відм} = p_3 = 0,18.$$

Відносна пропускна здатність ОЦ

$$q = 1 - P_{відм} = 1 - 0,18 = 0,82.$$

Тобто 18% заявок залишають систему не обслугованими.

Абсолютна пропускна здатність ОЦ

$$A = \lambda q = 1 * 0,82 = 0,82.$$

Середнє число зайнятих каналів - ПЕОМ

$$\bar{k} = \alpha(1 - P_{відм}) = 1,8 * (1 - 0,18) = 1,476.$$

Таким чином, при сталому режимі роботи СМО в середньому буде зайнято 1,5 комп'ютера з трьох - інші півтора простоюватимуть. Очевидно, що пропускну здатність ОЦ при даних λ і μ можна збільшити тільки за рахунок збільшення числа ПЕОМ.

Визначимо, скільки потрібно використовувати ПЕОМ, щоб скоротити імовірність відмови в обслуговуванні заявки в 2 рази, тобто щоб імовірність відмови у вирішенні задач не перевищувала 0,09. Для цього скористаємось формулою (5.23)

$$P_{відм} = \frac{(\alpha)^n}{n!} * p_0.$$

Результати розрахунків зведемо в таблицю 5.1.

Таблиця 5.1 – Розрахунок кількості ПЕОМ

n	1	2	3	4	5
$\mathbf{p_0}$	0,357	0,266	0,186	0,172	0,167
$\mathbf{P_{відм}}$	0,643	0,367	0,18	0,075	0,026

Таким чином, щоб зменшити імовірність відмови в обслуговуванні заявки в 2 рази необхідно додати ще один комп'ютер.

Контрольні запитання

1. Які задачі вирішує теорія масового обслуговування?
2. Перелічіть основні компоненти системи масового обслуговування, що зумовлюють її математичний опис.
3. Поясніть класифікацію систем масового обслуговування.
4. В якому випадку випадковий процес, що протікає в системі, називається марківським?
5. Дайте визначення стаціонарного потоку подій і регулярного потоку подій.
6. Які змінні є невідомими в рівняннях Колмогорова?
7. Поясніть, що таке фінальної імовірності?
8. Поясніть характеристики вхідного потоку вимог.
9. Якими параметрами визначаються потік заявок та потік обслуговувань?
10. Яким правилам може відповідати дисципліна черги в СМО?

ТЕМА 6
ЗАДАЧІ З УМОВАМИ НЕВИЗНАЧЕНОСТІ ТА КОНФЛІКТУ

6.1 Загальна математична постановка задачі стохастичного програмування

Економічні системи функціонують в умовах невизначеності, що зумовлює ризикованість прийнятих рішень. Прийняття рішень в умовах невизначеності та ризику вимагає визначення альтернативних дій, причому, досліджуване явище може характеризуватися заданим імовірнісним розподілом деяких параметрів. У такому випадку для прийняття рішень використовують методи ***стохастичного програмування***. Суть цих методів полягає в тому, що рішення залежить не тільки від керованих змінних $X=(x_1, x_2, \ldots, x_n)$, а так само від ряду ***випадкових*** некерованих параметрів $\omega=(\omega_1, \omega_2, \ldots, \omega_k)$. Предметом стохастичного програмування є екстремальні задачі, у яких параметри умов або складові розв'язку – ***випадкові величини***. Особливістю задач стохастичного програмування є те, що основні труднощі виникають не стільки під час розробки методів їх розв'язання, скільки під час постановки задачі. Це пов’язано з тим, що постановка задачі істотно залежить від структури наявної інформації й повинна відбивати особливості ухвалення рішення в умовах невизначеності. Зокрема, якщо вектор X є детермінованим, то він не залежить від випадкових параметрів моделі, а інакше потрібно враховувати його залежність від випадкових параметрів. Важливо також, що розуміють під максимізацією (мінімізацією) цільової функції: максимізацію її абсолютних значень або максимізацію її математичного сподівання або будь-якої іншої імовірнісної характеристики (наприклад, моди або середнього квадратичного відхилення). Треба вирішити так само, як повинні виконуватися обмеження: абсолютно для всіх ω або в середньому, або з малою ймовірністю припускається порушення обмежень задачі.

Таким чином, постановка задачі стохастичного програмування вимагає урахування економічних особливостей досліджуваного явища, а також евристичного підходу до її формулювання.

При виборі цільової функції в задачах стохастичного програмування треба визначити, чи цікавить нас, у першу чергу, максимум математичного сподівання економічного критерію (наприклад, прибутку або рентабельності), або мінімум його дисперсії. Також як цільову функцію можна прийняти лінійну комбінацію математичного сподівання та дисперсії економічного показника або ймовірність перевищення ним певного фіксованого значення. Помітимо, що для стохастичних задач актуальною є багатокритеріальна оптимізація.

Щодо обмежень, то вимога, щоб оптимальний розв'язок задовольняв їм за будь-яких значень випадкових параметрів ω, є занадто жорстким. Зазвичай припускають невиконання обмежень з певною досить малою ймовірністю α. Таким чином, задачі стохастичного програмування мають вигляд

знайти $$\max M[f(x, \omega)] \tag{6.1}$$

за обмежень

$$P\{g(x, \omega) \le 0\} \ge 1-\alpha. \tag{6.2}$$

У задачі (6.1)-(6.2) необхідно максимізувати середній сподіваний ефект за умови, що обмеження виконуються з імовірністю 1-α.

Або інакше

знайти $$\max \xi \tag{6.3}$$

за обмежень

$$P\{f(x, \omega) \ge \xi,\ g(x, \omega) \le 0\} \ge 1-\alpha. \tag{6.4}$$

У задачі (6.3)-(6.4) додаткова вимога, щоб значення критерію ефективності було не менш ξ з імовірністю 1-α, а значення ξ було максимальним.

Оптимальні плани, отримані на підставі моделей (6.1)-(6.2) або (6.3)-(6.4), називають **М-планами**, тому що критерієм оптимальності є математичне сподівання $f(x, \omega)$. Якщо як критерій прийнято дисперсію функції $f(x, \omega)$, то оптимальний план називається **D-планом**. Іноді як критерій оптимальності приймають різницю $M[f(x, \omega)]$-KD$[f(x, \omega)]$, де K - відомий параметр.

6.2 Класифікація задач стохастичного програмування. Основні методи розв'язання

Постановки задач стохастичного програмування різняться за трьома ознаками: за характером рішень; за показником якості розв'язку; за способом розчленовування обмежень задачі.

Задачі з цільовою функцією виду M[C X] називають **M-моделями**, задачі, в яких потрібно мінімізувати величину дисперсії, називають **D-моделями**, а стохастичні задачі, у яких максимізується ймовірність P, прийнято називати **P-моделями**. До цієї ж групи моделей включають також і задачі, де потрібно мінімізувати або максимізувати поріг ξ.

До **одноетапних** задач стохастичного програмування належать задачі, в яких рішення приймаються на підставі відомих стохастичних характеристик розподілу випадкових параметрів умов задачі, що отримані до спостереження за реалізаціями поточних значень цих параметрів. При цьому повинне прийматися певне найкраще в середньостатистичному значенні рішення.

Постановка задач у стохастичному програмуванні істотно залежить від можливості **уточнення** стану економічного середовища при реалізації оптимального рішення. Для економічних систем розробляють стратегічні і тактичні плани. У стратегічних планах враховують всі можливі значення ω. У певний момент часу в результаті спостереження стан економічного середовища стає відомим, тоді розробляють тактичний план, тобто знаходять розв'язок $X(\omega)$ при заданому ω. У загальному випадку спостереження не повністю визначають стан економічного середовища, тому етапи прийняття рішень можуть чергуватися з етапами спостережень. У цьому випадку виникає **N-етапна** задача стохастичного програмування. Властивість плану адаптуватися до умов його реалізації, що постійно змінюються, є необхідним для ефективного розвитку і функціонування економічної системи. Програмну частину плану вибирають таким чином, щоб максимізувати сподівану корисність з урахуванням майбутньої адаптації. План-адаптація має бути оптимальним для конкретної економічної ситуації і при цьому зберігати стратегію розвитку системи в цілому.

Методи розв'язання стохастичних задач поділяють на дві групи - **прямі** та **непрямі**. Прямі методи використовують, якщо на підставі інформації про параметр ω можна побудувати функції $f(x, \omega)$ і $g_i(x, \omega)$. У непрямих методах стохастичну задачу приводять до задачі лінійного або нелінійного програмування, тобто розглядають детермінований аналог задачі стохастичного програмування.

Зокрема, **імовірнісне динамічне програмування** відрізняється від детермінованого тим, що стани та прибутки на кожному етапі є випадковими величинами. Моделі імовірнісного динамічного програмування виникають, наприклад, під час розглядання стохастичних моделей управління запасами. **Одноетапні моделі** управління запасами відбивають ситуацію, коли для задоволення попиту протягом певного періоду продукція замовляється тільки один раз. **Багатоетапні моделі** використовують під час планування запасів на n періодів. Вони припускають, що попит у кожний період описується стаціонарною (незалежною від часу) щільністю ймовірностей.

Апарат теорії ігор розрізняє стратегічні та статистичні ігри. В основі стратегічних ігор лежить припущення, що кожен з гравців діє активно, і їхні інтереси є протилежними. У теорії статистичних ігор одним з гравців є природа, тобто сукупність зовнішніх обставин, в умовах яких доводиться приймати рішення. Неминучою платою за спробу одержати рішення в умовах неповної інформації про закони природи є прийняття помилкових рішень. Теорія статистичних ігор дозволяє виробити таку стратегію щодо прийняття рішень, що хоча й не виключає можливості прийняття невірних рішень, але зводить до мінімуму пов'язані з цим небажані наслідки.

Приклад 6.1. Нехай необхідно зробити запас n товарів у кількості $X=(x_1, x_2, \ldots, x_n)$, на які є випадковий попит $\omega=(\omega_1, \omega_2, \ldots, \omega_n)$. Нестача одиниці j-го товару обкладається штрафом c_j, тобто $C=(c_1, c_2, \ldots, c_n)$, а витрати на зберігання одиниці відповідного товару, що не вдалося збути, задані вектором $d=(d_1, d_2, \ldots, d_n)$. Функція збитку, що відповідає розв'язку x, має вигляд

$$f(x,\omega)=\sum_{j=1}^{n}\left\{Cj\max\left(0,\omega_j-x_j\right)+d_j\max\left(0,x_j-\omega_j\right)\right\},$$

де $Cj\max\left(0,\omega_j-x_j\right)$ - штраф за незадоволення попиту на j-й вид товару;

$d_j\max\left(0,x_j-\omega_j\right)$ - витрати на зберігання j-го виду товару.

Для знаходження оптимального розв'язку даної задачі необхідно знати функцію розподілу випадкової величини ω. Оскільки така функція невідома, уважають, що ω розподілена рівномірно. При цьому необхідно враховувати, що це припущення може призвести до ухвалення неправильного рішення.

6.3 Імітаційне моделювання

Методи імітаційного моделювання дозволяють зібрати необхідну інформацію про поведінку системи шляхом створення її комп’ютеризованої моделі. Імітаційне моделювання не вирішує оптимізаційних задач, а являє собою техніку оцінки значень функціональних характеристик системи, що моделюється. Методи імітаційного моделювання знаходять широке застосування в економічних і комерційних задачах, включаючи оцінку поведінки споживача, визначення цін, економічне прогнозування діяльності фірм, у соціальних задачах та ін.

Попередником сучасного імітаційного моделювання вважається метод Монте-Карло, основна ідея якого полягає у використанні вибірки випадкових чисел для одержання імовірнісних або детермінованих оцінок будь-яких величин. Імітація є випадковим експериментом, тому будь-який результат імітаційного моделювання піддається експериментальним помилкам і підлягає статистичній перевірці. Для будь-якого експерименту так само важливим є питання, яким має бути обсяг вибірки n і число реалізацій досліджуваної випадкової величини N.

Відмінність сучасних імітаційних моделей від методу Монте-Карло полягає в тому, що імітаційна модель зазвичай пов'язана з вивченням реально існуючої системи, поведінка якої є функцією часу. Існує два типи імітаційних моделей. **Безперервні моделі** використовуються для систем, поведінка яких змінюється в часі безупинно. Безперервні імітаційні моделі зазвичай представляються у вигляді різницево-диференційних рівнянь, які описують взаємодію між різними елементами системи. **Дискретні моделі** описують системи, поведінка яких змінюється тільки в певні моменти часу. Типовим прикладом такої моделі СМО, що являє собою систему, зміни в якій відбуваються лише тоді, коли клієнт надходить у чергу або залишає систему після обслуговування. Це означає, що в будь-якій дискретній імітаційній моделі є дві головних події, за якими необхідно досліджувати систему. В імітаційній моделі події, пов'язані з прибуттям, визначаються часом між надходженнями клієнтів, а події, пов'язані з їхнім доглядом, - часом обслуговування.

Випадковість в імітаційних моделях виникає тоді, коли інтервал часу t між однорідними подіями є випадковим. Відомий ряд методів одержання послідовних випадкових значень $t=t_1, t_2, \dots$, що мають заданий розподіл імовірностей $f(x)$. Розглянемо два з них: метод зворотних функцій і метод згорток.

Обидва методи ґрунтуються на використанні незалежних однаково розподілених випадкових чисел, що мають рівномірний розподіл на інтервалі [0,1]. Метод зворотних функцій використовується для безперервних розподілів, наприклад для експонентного або рівномірного. Метод згорток використовується в складніших ситуаціях, наприклад, під час генерації випадкових чисел, що мають нормальний розподіл або розподіл Пуассона.

Метод зворотних функцій вимагає виконання наступних дій. Спочатку генерується випадкове число R з інтервалу [0,1], потім обчислюється шукане випадкове число $x=F^{-1}(R)$, де F^{-1}– функція, зворотна до функції розподілу $F(x)=P\{X<x\}$ (рис. 6.1). Метод заснований на тому, що якщо функцію розподілу $F(x)$ розглядати як випадкову величину, то вона розподілена рівномірно на інтервалі [0,1].

Приклад 6.2. Нехай час появи клієнтів розподілений за експонентним законом з параметром λ=4. Функція розподілу має вигляд $F(t)=1-e^{-\lambda t}=1-e^{-4t}$.

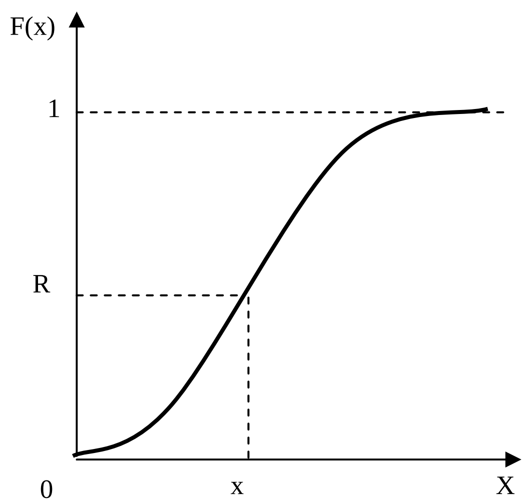


Рисунок 6.1 – Метод зворотних функцій

Урахуємо, що $F(t)=R$, одержимо $t=\frac{1}{\lambda}\ln(1-R)$. Оскільки R - випадкове число з інтервалу [0,1] і (1-R) так само випадкове число з того самого інтервалу, можна замінити (1-R) на R. Нехай R=0,9, тоді одержимо одне конкретне значення інтервалу часу між клієнтами $t=\frac{1}{4}\ln(1-0{,}9)=0{,}577$.

Значення R мають обиратися випадково з інтервалу [0,1] і підпорядковуватися рівномірному розподілу.

Основна ідея **методу згорток** полягає в тому, щоб виразити шукану випадкову величину у вигляді суми інших випадкових величин, для яких легко отримати реалізації випадкових значень. Для одержання значень, що відповідають нормальному розподілу з математичним сподіванням m і стандартним відхиленням σ використовують центральну граничну теорему. Нагадаємо, що суть її збігається до того, що сума n однаково розподілених випадкових величин прагне до нормального розподілу при нескінченому збільшенні n. Нехай $x=R_1+R_2+\ldots+R_n$, де R_1, R_2, …, R_n – випадкові числа, рівномірно розподілені в інтервалі [0,1]. Відповідно до центральної граничної теореми випадкова величина x є асимптотично нормальною величиною із середнім $n/2$ і дисперсією $n/12$. Тоді випадкова величина X з математичним сподіванням m і стандартним відхиленням σ визначається за формулою

$$X = m + \sigma\left(\frac{x - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right).$$

У практичних розрахунках для зручності зазвичай приймають n=12, тоді $X = m + \sigma(x-6)$.

Імітаційне моделювання являє собою статистичний експеримент. Його результати повинні ґрунтуватися на відповідних статистичних перевірках з використанням, наприклад, довірчих інтервалів і методів перевірки гіпотез. Для цього спостереження повинні задовольняти наступним вимогам: мати стаціонарні розподіли, тобто, що не змінюються під час проведення експерименту, підпорядковуватися нормальному розподілу та бути незалежними.

Характер імітаційних обчислень стимулює створення спеціалізованих мов програмування. У цей час на ринку програмних продуктів для моделювання домінують комерційні пакети Arena, AweSim і GPSS/H, які мають розвинений інтерфейс, що спрощує процес створення імітаційних моделей.

6.4 Прийняття рішень в умовах невизначеності та ризику

У теорії прийняття рішень використовують процедури вибору найкращої з кількох можливих альтернатив. Наскільки правильним буде вибір, залежить від якості даних, використовуваних при описі ситуації, у якій приймається рішення. З цього погляду процес прийняття рішень може належати до однієї з трьох можливих умов. Прийняття рішень в умовах визначеності, коли дані відомі точно; прийняття рішень в умовах ризику, коли дані можна описати за допомогою імовірнісних розподілів; прийняття рішень в умовах невизначеності, коли даним не можна приписати вагові коефіцієнти, які б зумовлювали ступінь їхньої значущості в процесі прийняття рішень.

Якщо рішення приймається в умовах ризику, вагові коефіцієнти альтернативних рішень описують імовірнісними розподілами. Для ухвалення рішення в цьому випадку використовують **критерій сподіваного значення**, відповідно до якого альтернативні рішення порівнюють з погляду максимізації сподіваного прибутку або мінімізації сподіваних витрат. При цьому припускається, що прибуток (витрати), пов'язаний з кожним альтернативним рішенням, є випадковою величиною. Розглянемо ситуацію, пов'язану з ухваленням рішення за наявності кінцевого числа альтернатив і точних значень матриці доходів.

Приклад 6.2. Необхідно вкласти на фондовій біржі 10 тис.дол. в акції однієї з двох компаній A або B. Акції компанії A є ризикованими, але можуть принести 50% прибутку від суми інвестиції протягом року. Якщо умови фондової біржі будуть несприятливі, сума інвестиції може знецінитися на 20%. Компанія B забезпечує безпеку інвестицій з 15% прибутку в умовах підвищення котирувань на біржі і тільки 5% в умовах зниження котирувань. Аналітичні публікації з імовірністю 60% прогнозують підвищення котирувань. У яку компанію краще вкласти гроші? Складемо таблицю 6.1.

Таблиця 6.1 – Вихідні дані

Альтернативні рішення	**Прибуток від інвестиції 10 тис. дол.**	
	При підвищенні котирувань	**При зниженні котирувань**
Акції компанії A	5000	-2000
Акції компанії B	1500	500
Імовірність	0,6	0,4

Розв'язання

Визначимо сподіваний прибуток. Для акцій компанії A

$$5000 \times 0{,}6 + (-2000) \times 0{,}4 = 2200 \text{ дол.}$$

Для акцій компанії B

$$1500 \times 0{,}6 + 500 \times 0{,}4 = 1100 \text{ дол.}$$

Таким чином, рішенням, заснованим на цих обчисленнях, є покупка акцій компанії A.

У цьому випадку підвищення та зниження котирувань на біржі є станами природи. У загальному випадку задача прийняття рішень може включати n станів природи і m альтернатив. Якщо p_j – імовірність j-го стану природи, а a_{ij} - платіж, пов'язаний з ухваленням рішення, то очікуваний платіж для рішення i обчислюється в такий спосіб

$$M[v_i] = a_{i1}p_1 + a_{i2}p_2 + ... + a_{in}p_n, \quad де \quad \sum p_j = 1.$$

Найкращим рішенням буде те, що відповідає $M[v_i]^* = \max\{M[v_i]\}$ або $M[v_i]^* = \min\{M[v_i]\}$ залежно від того, чи є платіж у задачі доходом або збитком.

Критерій сподіваного значення має дві модифікації, перша з яких полягає у визначенні **апостеріорних імовірностей** на основі експерименту над досліджуваною системою, а друга – у визначенні **функції корисності**.

Імовірності, які використовуються для формулювання критерію сподіваного значення, визначаються, як правило, на підставі попередньо накопиченої інформації. Іноді виявляється можливим перерахувати ці ймовірності за допомогою поточної інформації, отриманої на підставі вибіркових (або експериментальних) даних. Одержувані при цьому ймовірності називають апостеріорними (або бейєсовськими) на відміну від апріорних, отриманих з вихідної інформації.

Функцію корисності U(x) використовують, коли важливіша не реальна величина платежів, а скоріше їхня корисність, зумовлена ставленням особи, що приймає рішення (ОПР), до ризику. Нехай ОПР має шанс 50 на 50, що інвестиція в 20 тис. дол. або принесе прибуток в 40 тис. дол. або буде цілком загублена. Різні індивідууми проявляють різне ставлення до ризику, тобто вони виявляють стосовно ризику різну корисність. У розглянутій ситуації найкращий платіж дорівнює 40 тис. дол., а найгірший - -20 тис. дол. Встановимо довільну шкалу корисності U – від 0 до 100. 0 відповідає корисності -20, а 100 – 40, тобто U(-20)=0; U(40) =100. Для визначення загального вигляду функції корисності визначимо корисність у точках між -20 і 40.

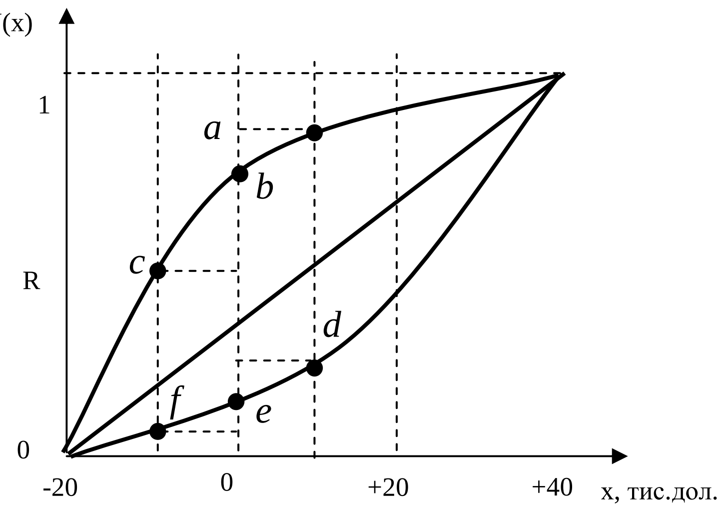


Рисунок 6.2 – Функція корисності

Якщо ставлення до ризику ОПР байдуже, то функція корисності є прямою, що з'єднує точки (-20, 0) і (40, 100), як показано на рисунку 6.2. У цьому випадку як реальні гроші, так і їхня корисність дають рішення, що збігаються. Нехай ОПР X не прихильне до ризику, тоді для нього при зміні в 10 тис. дол. вправо та уліво від 0 збільшення прибутку змінює корисність на величини *ab* і *bc* відповідно, причому, *ab*<*bc*. До таких самих змін ОПР *Y* ставиться інакше, у цьому випадку *de*>*ef*.

Визначимо корисність, що відповідає проміжним значенням -10; 0; 10; 20. Функція корисності обчислюється за формулою

$$U(x)=p(-20)+(1-p)\mathrm{U}(40).$$

Для визначення значення U(x) просять ОПР повідомити свою перевагу між гарантованою наявною сумою x і можливістю з імовірністю р програти 20 тис. дол. або з імовірністю (1-p) виграти 40 тис. дол. Під перевагою тут розуміють вибір такої ймовірності p, за якою пропоновані варіанти однаково привабливі. Наприклад, при x=20 тис. дол. ОПР повідомляє, що p=0,8, тоді U(20) =100-100*0,8=20. Аналогічно знаходять ряд точок для визначення форми функції корисності, а потім визначають шукану криву за допомогою регресійного аналізу або лінійної інтерполяції.

Прийняття рішень в **умовах невизначеності**, як і в умовах ризику, вимагає визначення альтернативних дій, яким відповідають «платежі», що залежать від **станів природи**. Запишемо матрицю платежів у задачі прийняття рішень з m можливими діями та n станами природи

$$
\begin{array}{c|cccc|}
 & \omega_1 & \omega_2 & \dots & \omega_n \\
\hline
a_1 & v(a_1, \omega_1) & v(a_1, \omega_2) & \dots & v(a_1, \omega_n) \\
a_2 & v(a_2, \omega_1) & v(a_2, \omega_2) & \dots & v(a_1, \omega_n) \\
\dots & \dots & \dots & \dots & \dots \\
a_m & v(a_m, \omega_1) & v(a_m, \omega_2) & \dots & v(a_m, \omega_n)
\end{array}
$$

Елемент a_i представляє і-й можливий розв'язок, а елемент ω_j - j-й стан природи. Плата або дохід, пов'язаний з розв'язком a_i і станом ω_j дорівнює $v(a_i, \omega_j)$.

Відмінність між прийняттям рішень в умовах ризику та невизначеності полягає в тому, що в умовах невизначеності імовірнісний розподіл, що відповідає станам ω_j невідомий. Ця обставина зумовила розвиток спеціальних критеріїв для аналізу ситуації, пов'язаної з прийняттям рішень: критерію Лапласа, мінімаксного критерію, критерію Севіджа та критерію Гурвіца. Ці критерії відрізняються за ступенем консерватизму особи, що приймає рішення.

Критерій Лапласа спирається на принцип недостатньої підстави, який говорить, що оскільки розподіл імовірностей станів $P\{\omega_j\}$ невідомий, немає підстав уважати їх різними. Отже, робиться припущення, що ймовірності всіх станів природи дорівнюють одна одній, тобто $P\{\omega_1\}= P\{\omega_2\}=\dots=P\{\omega_n\}=1/n$. Якщо при цьому $v(a_i, \omega_j)$ є прибутком, то найкращим рішенням є те, що забезпечує

$$\max_{ai}\left\{\frac{1}{n}\sum_{j=1}^{n} v(a_i, \omega_j)\right\}.$$

Якщо величина $v(a_i, \omega_j)$ являє собою видатки, то максимум замінюють на мінімум.

Максимінний (мінімаксний) критерій заснований на консервативній обережні поведінцці особи, що приймає рішення, і збігається до вибору найкращої альтернативи з найгірших. Якщо величина $v(a_i, \omega_j)$ є прибутком, то відповідно до максимінного критерію за оптимальне обирається рішення, що забезпечує

$$\max_{ai}\left\{\min_{\omega j} v(a_i, \omega_j)\right\}.$$

Якщо величина $v(a_i, \omega_j)$ являє собою втрати, використовується мінімаксний критерій

$$\min_{ai}\left\{\max_{\omega j} v(a_i, \omega_j)\right\}.$$

Критерій Севіджа знижує консерватизм мінімаксного (максимінного) критерію шляхом заміни матриці платежів $v(a_i, \omega_j)$ матрицею втрат $r(a_i, \omega_j)$, що визначається в такий спосіб

$$r(a_i, \omega_j) = \begin{cases} \max_{ak}\left\{v(a_k, \omega_j)\right\} - v(a_i, \omega_j), & \text{якщо } v - \text{прибуток}, \\ v(a_ш, \omega_j) - \min_{ak}\left\{v(a_k, \omega_j)\right\}, & \text{якщо } v - \text{втрати}. \end{cases}.$$

Приклад 6.3. Є матриця платежів

	ω_1	ω_2
a_1	11000	90
a_2	10000	10000

Максимум рядка 1 становить 11000, а рядка 2 – 10000 (мінімакс). Застосування мінімаксного критерію призводить до того, що рішення a_2 з фіксованими втратами 10000 є кращим. Однак можна вибрати і a_1, тому що в цьому випадку існує можливість втратити лише 90, якщо реалізується стан ω_2 при потенційному виграші 11000.

Визначимо, який результат вийде, якщо в мінімаксному критерії замість матриці платежів $v(a_i, \omega_j)$ скористатися матрицею втрат $r(a_i, \omega_j)$.

	ω_1	ω_2
a_1	1000	0
a_2	0	9910

Максимум рядка 1 становить 1000 (мінімакс), а рядка 2 – 9910. Як бачимо, мінімаксний критерій, застосований до матриці втрат, призводить до вибору рішення a_1.

Критерій Гурвіца охоплює ряд різних підходів до прийняття рішень – від оптимістичнішого до найпесимістичнішого. Нехай $0 \leq \lambda \leq 1$ і величини $v(a_i, \omega_j)$ є доходами. Тоді рішенню, обраному за критерієм Гурвіца, відповідає

$$\max_{ai}\left\{\lambda \max_{\omega j} v(a_i, \omega_j) + (1-\lambda) \min_{\omega j} v(a_i, \omega_j)\right\}.$$

Параметр λ - показник оптимізму. Якщо $\lambda=0$, то критерій Гурвіца стає еквівалентним застосуванню мінімаксного критерію. Якщо $\lambda=1$, критерій Гурвіца стає занадто оптимістичним, тому що розраховує на найкраще з найкращих рішень. Можна конкретизувати ступінь оптимізму вибором величини λ з інтервалу [0, 1]. Найрозумнішим уявляється вибір $\lambda=0,5$.

Якщо величини $v(a_i, \omega_j)$ є втратами, то критерій приймає наступний вигляд

$$\min_{ai}\left\{\lambda \min_{\omega j} v(a_i, \omega_j) + (1-\lambda) \max_{\omega j} v(a_i, \omega_j)\right\}.$$

Приклад 6.4. Передбачається побудова готеля в одному з регіонів України на Полтавщині, у Каневі, на Херсонщині або у Прикарпатті. Прибуток залежить від різних факторів: місцевих умов, клімату, наявності водойму, якості доріг та ін. Для кожного регіону виділено чотири рівні прибутку, що враховують вплив факторів та наведені у вигляді матриці

$$P = \begin{vmatrix} 10 & 4 & 15 & 7 \\ 15 & 10 & 7 & 19 \\ 4 & 6 & 7 & 22 \\ 2 & 8 & 8 & 16 \end{vmatrix}.$$

Рядки матриці відповідають регіонам. Скористаємось максимінним критерієм (критерієм Вальда)

$$\max_{ai}\left\{\min_{\omega j}(a_{i\,j})\right\} = \max_{i}\{4,7,4,2\} = 7,$$

отже, максимальний з мінімальних прибуток відповідає i=2, тобто другому рядку (Канєв).

Для використання критерію Севіджа замінимо матрицю прибутків $v(a_i, \omega_j)$ матрицею ризиків $r(a_i, \omega_j)$, що визначається в такий спосіб

$$r(a_i, \omega_j) = \begin{cases} \max\limits_{ak}\left\{v(a_k, \omega_j)\right\} - v(a_i, \omega_j), & \textit{якщо } v - \textit{прибуток}, \\ v(a_ш, \omega_j) - \min\limits_{ak}\left\{v(a_k, \omega_j)\right\}, & \textit{якщо } v - \textit{втрати}. \end{cases}.$$

Для визначення елементів матриці ризиків в кожному стовпчику обирають максимальний елемент та з нього віднімають всі елементи стовпчика:

$$R = \begin{vmatrix} 5 & 6 & 0 & 15 \\ 0 & 0 & 8 & 3 \\ 11 & 4 & 8 & 0 \\ 13 & 2 & 7 & 6 \end{vmatrix}.$$

Обчислимо $\min\limits_{ai}\left\{\max\limits_{\omega j}(r_{i\,j})\right\} = \min\limits_{i}\{15,8,11,13\} = 8$, що так само відповідає другому рядку i=2.

Скористаємось критерієм Гурвіца. Нехай λ=0,5, тоді рішенню, обраному за критерієм Гурвіца, відповідає

$$\max_{ai}\left\{0{,}5\max_{\omega j} v(a_i, \omega_j) + 0{,}5\min_{\omega j} v(a_i, \omega_j)\right\} = \max_{ai}\left\{9{,}5;13;13;9\right\} = 13,$$

де елементи обчислюємо за рядками. Отже найкращі рішення за критерієм Гурвіца відповідають другому та третьому рядкам (i=2 або i=3).

За критерієм Лапласа найкращим рішенням є те, що забезпечує

$$\max_{ai}\left\{\frac{1}{n}\sum_{j=1}^{n} v(a_i, \omega_j)\right\},$$

тобто потрібно обрати найбільше математичне сподівання прибутку за варіантами рішення

$$\max_{ai}\left\{\frac{10+4+15+7}{4}; \frac{15+10+7+19}{4}; \frac{4+6+7+22}{4}; \frac{2+8+8+16}{4}\right\} =,$$
$$= \max_{ai}\left\{9;\ 12{,}75;\ 9{,}5;\ 8{,}5\right\} = 12{,}75.$$

що знову відповідає i=2.

Отже аналіз ситуації за критеріями дає підставу обрати Канєв.

Контрольні запитання

1. Поясніть суть задач стохастичного програмування.
2. Яка стохастична задача називається одноетапною? Багатоетапною? Що розуміють під етапами?
3. Дайте загальну характеристику методів стохастичного програмування.
4. Які класи задач стохастичного програмування прийнято розрізняти?
5. Поясніть розходження між стратегічною та статистичною іграми.
6. В яких випадках використовують метод імітаційного моделювання? У чому відмінність безперервних і дискретних імітаційних моделей?
7. Для чого призначений метод зворотних функцій? Поясніть сутність методу.
8. Для чого призначений метод згорток? Поясніть сутність методу.
9. Яка відмінність між прийняттям рішень в умовах визначеності, в умовах невизначеності та в умовах ризику?
10. У яких випадках для ухвалення рішення використовують критерій сподіваного значення?

ТЕМА 7
ЗАДАЧІ ТЕОРІЇ ІГОР

7.1 Основні поняття та визначення

На практиці часто зустрічаються завдання, коли рішення приймаються в умовах невизначеності, у конфліктних ситуаціях (відносини споживача та продавця, банку та клієнта, гравців у шахи, шашки та ін.). Методи теорії ігор дозволяють знаходити оптимальні рішення в умовах конфлікту. Математична модель конфліктної ситуації називається грою, її учасники - гравцями. Для кожної гри вводяться правила, що визначають варіанти дій гравців, обсяг інформації про поведінку партнерів (супротивників), виграш у результаті гри.

Ігри можуть бути класифіковані за різними ознаками. Насамперед, вони можуть відрізнятися числом гравців. Ігри двох гравців, до яких збігаються багато ігор з великим числом учасників, найпоширеніші і гарно вивчені. Парні ігри широко використовуються в практичних застосуваннях і будуть розглянуті нижче.

За характером виграшів ігри розділяються на ігри з нульовою сумою та з ненульовою. При грі з нульовою сумою сумарний виграш всіх гравців дорівнює нулю, тобто загальний капітал перерозподіляється між гравцями. Парна гра з нульовою сумою називається антагоністичною (виграш одного гравця дорівнює програшу іншого).

За кількістю стратегій ігри розділяються на кінцеві та нескінченні, якщо хоча б один гравець має нескінченне число стратегій (наборів правил на вибір ходу гри).

За видом функцій виграшу ігри можуть бути матричними, біматричними, безперервними, опуклими, типу дуелей та ін. Досить добре вивчена матрична гра - парна гра з нульовою сумою, у якій виграші першого гравця (програші другого) задаються у вигляді матриці. Будь-яка матрична гра має розв'язання та збігається до завдання лінійного програмування.

На функції виграшу можуть бути накладені вимоги безперервності та опуклості.

За кількістю ходів ігри розділяються на однокрокові та багатокрокові, які у свою чергу мають низку різновидів.

Якщо гравці мають можливість поєднуватися та діяти спільно, то гра стає коаліційною.

Залежно від використання інформації розрізняють ігри з повною інформацією (кожному гравцеві відомі ходи партнерів) і з неповною інформацією, коли гравцям відомі не всі ходи партнерів.

Теорія ігор визначає оптимальні стратегії для кожного гравця, дотримуючись яких один гравець повинен одержати максимальний виграш, а другий - мінімальний програш. Оптимальна стратегія повинна мати властивість стійкості, тобто кожному з гравців має бути невигідно відмовитися від своєї стратегії.

7.2 Ціни антагоністичних ігор. Сідлові точки, їхні властивості

Оптимальні стратегії матричних ігор безпосередньо зв'язані з властивостями сідлових точок, які відіграють важливу роль у теорії оптимізації і зокрема під час розв’язання завдань нелінійної оптимізації.

В антагоністичній (парний з нульовою сумою) грі перший гравець має множину стратегій А, другий - множину стратегій В; функція $F(x,y)$ визначає виграш першого гравця в ситуації (x,y), коли перший гравець використовує стратегію $x \in A$, другий - $y \in B$. Стратегії x і y називаються чистими стратегіями гравців.

В антагоністичній грі перший гравець з мінімальних виграшів, які забезпечує другий гравець, знаючи стратегію першого, прагне за можливістю одержати максимальний виграш α, тобто

$$\alpha = \max_{x \in A} \min_{y \in B} F(x, y) . \qquad (7.1)$$

Значення α називається нижньою ціною гри й показує, який мінімальний виграш гарантує собі перший гравець, застосовуючи чисту максимінну стратегію.

Другий гравець прагне максимально зменшити виграш першого гравця. Тому він визначає максимально можливий для кожної стратегії першого гравця виграш $\max_{x \in A} F(x, y)$ і обирає стратегію, що його мінімізує, тобто другий гравець не дозволяє одержати першому виграш, що перевищує β (мінімаксна стратегія):

$$\beta = \min_{y \in B} \max_{x \in A} F(x, y) . \qquad (7.2)$$

Величина β називається верхньою ціною гри.

Між цінами має місце співвідношення, що виражається лемою 7.1.

Лема 7.1. Величини α і β, що визначаються за виразами (7.1) і (7.2), задовольняють нерівності

$$\alpha \le \beta, \quad \textit{тобто} \quad \max_x \min_y F(x,y) \le \min_y \max_x F(x,y). \qquad (7.3)$$

Доказ випливає з визначення максимуму та мінімуму функцій:

$$\min_y F(x,y) \le F(x,y) \le \max_x F(x,y),$$

або

$$\min_y F(x,y) \le \max_x F(x,y). \qquad (7.4)$$

У лівій частині нерівності (7.4) x приймає будь-яке значення $x \in A$, тому

$$\max_x \min_y F(x,y) \le \max_x F(x,y). \qquad (7.5)$$

У праву частину нерівності (7.5) входить кожне $y \in B$, тому

$$\max_x \min_y F(x,y) \le \min_y \max_x F(x,y).$$

Таким чином, якщо обидва гравця поводяться розумно, виграш першого лежить у межах $\alpha < F(x,y) < \beta$. Якщо $\alpha = \beta = \nu$, то говорять, що ціна гри дорівнює ν, стратегії x_0, y_0 гравців є оптимальними, а функція $F(x,y)$ має сідлову точку (x_0,y_0).

Сідловою точкою (x_0,y_0) дійсної функції двох змінних $F(x,y)$, $x \in A$, $y \in B$, називається пара (x_0,y_0), для якої виконуються співвідношення

$$F(x,y_0) \le F(x_0,y_0) \le F(x_0,y),$$

тобто

$$\max_x F(x,y_0) = F(x_0,y_0) = \min_y F(x_0,y). \qquad (7.6)$$

З визначення сідлової точки випливає, що якщо один з гравців дотримується оптимальної стратегії, то іншому найбільше вигідно також дотримуватися цієї стратегії.

Має місце важливий зв'язок між існуванням ціни гри та сідлової точки.

Теорема 7.1. Нехай для функції $F(x,y)$, $x \in A$, $y \in B$, існують $\max_x \min_y F(x,y)$ і $\min_y \max_x F(x,y)$. Тоді необхідною й достатньою умовою виконання рівності

$$\max_x \min_y F(x,y) = \min_y \max_x F(x,y) \qquad (7.7)$$

є існування сідлової точки (x_0,y_0) функції $F(x,y)$, причому

$$F(x_0,y_0) = \max_x \min_y F(x,y) = \min_y \max_x F(x,y). \qquad (7.8)$$

Доказ.

Необхідність. З умови (7.7) випливає існування таких $x_0 \in A$, $y_0 \in B$, що

$$\max_x \min_y F(x,y) = \min_y F(x_0,y),$$

$$\min_y \max_x F(x,y) = \max_x F(x_0,y),$$

причому

$$\min_y F(x_0,y) = \max_x F(x,y_0).$$

За визначенням максимуму та мінімуму

$$\min_y F(x_0, y) \le F(x0, y0) \le \max_x F(x, y_0) = \min_y F(x_0, y),$$

тобто виконується співвідношення (7.6), і (x_0,y_0) є сідлова точка, причому

$$F(x_0, y_0) = \min_y F(x_0, y) = \max_x \min_y F(x, y) = \max_x F(x, y_0) = \min_y \max_x F(x, y),$$

Достатність. Нехай (x_0,y_0) - сідлова точка. За визначенням мінімуму та максимуму для будь-яких точок, у тому числі (x_0,y_0), справедливі нерівності

$$\min_y \max_x F(x, y) \le \max_x F(x, y_0), \ \max_x \min_y F(x, y) \ge \min_y F(x_0, y). \qquad (7.9)$$

З нерівностей (7.9) і співвідношення (7.6) випливає

$$\min_y \max_x F(x, y) \le \max_x \min_y F(x, y).$$

З урахуванням зворотної нерівності (7.3)

$$\min_y \max_x F(x, y) = \max_x \min_y F(x, y).$$

Якщо α<β, то ціни гри ν у чистих стратегіях не існує і необхідно використовувати змішані стратегії, серед яких шукаються оптимальні рішення. Для кінцевої множини A чистих стратегій змішаною стратегією називається вектор p = (p_1, p_2, …, p_n), де p_i - імовірність вибору чистої стратегії x_i, $p_i \ge 0$, $\sum_{i=1}^{n} p_i = 1, \quad i = \overline{1,n}$.

Очевидно, чиста стратегія x_i є окремим випадком змішаної, що задається вектором p = (0, 0, ..., 1, 0,... 0), і-а компонента якого дорівнює одиниці, а всі інші - нулю.

Розглянемо знаходження оптимальних стратегій для матричних антагоністичних ігор.

7.3 Матричні ігри в чистих та змішаних стратегіях

Нехай множини стратегій $x \in$ A і $y \in$ B містять кінцеве число точок; поклавши x=i, y=j, $F(x,y)=F(\text{i,j})=a_{\text{ij}}$, гру можна задати платіжною матрицею P:

$$P = (a_{ij}) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{vmatrix}.$$

Якщо перший гравець обирає i-у чисту стратегію (i-й рядок), $i = \overline{1,n}$, а другий j-у стратегію (j-й стовпець), $j = \overline{1,m}$, то виграш першого гравця в ситуації (i,j) дорівнює a_{ij}. Нижньою та верхньою цінами матричної гри в чистих стратегіях відповідно до рівностей (7.1), (7.2) буде

$$\alpha = \max_{ш} \min_{о} a_{ij}, \ \beta = \min_j \max_i a_{ij}. \qquad (7.10)$$

Якщо α=β, матриця P має сідлову точку (i_0,j_0) і для будь-яких i, j (з урахуванням співвідношення (7.6)) виконуються нерівності

$$a_{ij0} \le a_{i0j0} \le a_{i0j}. \qquad (7.11)$$

Ціна гри $v = a_{i0\,j0}$, стратегії i_0, j_0 є оптимальними.

Зауваження 1. Зі співвідношення (7.11) випливає простий спосіб відшукання сідлової точки: у кожному рядку знаходять мінімальний елемент і перевіряють, чи буде він максимальним у своєму стовпці. Якщо він є таким, то представляє сідлову точку.

Приклад 7.1. Визначити, чи існує ціна гри в чистих стратегіях для гри, що задана матрицею Р:

$$\begin{array}{cc} & & \alpha_i \\ P= & \left|\begin{array}{ccc} 0,7 & 0,6 & 0,9 \\ 0,8 & 0,5 & 0,8 \\ 0,8 & 0,6 & 0,7 \end{array}\right| & \begin{array}{c} \underline{0,6} \\ 0,5 \\ \underline{0,6} \end{array} \\ \beta_j & \begin{array}{ccc} 0,8 & \underline{0,6} & 0,9 \end{array} & \end{array}$$

За рядками обираємо мінімальні елементи $\alpha_i = \min\limits_{o} a_{ij}$, серед них знаходимо максимальні $\alpha_1 = \alpha_3 = 0,6$.

Серед максимальних елементів за стовпцями $\beta_j = \max\limits_{i} a_{ij}$ виділяємо мінімальний $\beta_2 = 0,6$.

Відповідно до умови (7.10) $\alpha=\beta=v=0,6$; гра має дві сідлові точки (1,2), (3,2), тобто в першого гравця дві оптимальні чисті стратегії (перша і третя), у другого - друга. В обох випадках перший гравець має виграш (другий - програш) 0,6.

Приклад 7.2. Підприємство випускає три види продукції, маючи прибуток, що залежить від попиту. У матриці Р елементи a_{ij} характеризують прибуток, одержуваний при реалізації продукції i-го виду у j-му стані попиту. Чи існує чиста оптимальна стратегія випуску продукції?

$$\begin{array}{ccc} & & \alpha_i \\ P= & \left|\begin{array}{cccc} 3 & 2 & 6 & 8 \\ 9 & 10 & 5 & 6 \\ 7 & 8 & 4 & 5 \end{array}\right| & \begin{array}{c} 2 \\ \underline{5} \\ 4 \end{array} \\ \beta_j & \begin{array}{cccc} 9 & 10 & \underline{6} & 8 \end{array} & \end{array}$$

Знаходимо $\alpha=5$, $\beta=6$. Оскільки $\alpha\neq\beta$, оптимальне рішення треба шукати в змішаних стратегіях.

Обсяг розрахунків можна скоротити, використавши принцип домінування. Якщо всі елементи i-го рядка більше або дорівнюють відповідним елементам *k*-го рядка, то i-й рядок домінує *k*-й і останній можна виключити з матриці Р, оскільки вона свідомо невигідна першому гравцеві. Аналогічно, якщо j-й стовпець домінує *s*-й, то j-а стратегія невигідна другому гравцеві й j-й стовпець може бути також виключений з матриці.

У прикладі 7.2 другий рядок домінує третій, четвертий стовпець домінує третій, тому третій рядок і четвертий стовпець можна викреслити з вихідної матриці P, одержавши простішу матрицю P_1.

$$\begin{array}{rccc|c} & & & & \alpha_i \\ P_1= & 3 & 2 & 6 & 2 \\ & 9 & 10 & 5 & \underline{5} \\ \beta_j & 9 & 10 & \underline{6} & \end{array}$$

Для матриці P_1 знову $\alpha = 5$, $\beta = 6$, $\alpha \neq \beta$.

Нехай перший гравець обирає змішану стратегію $p = (p_1, p_2, \ldots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1, \quad i = \overline{1,n}$, де p_i - імовірність вибору i-ої стратегії, а другий – стратегію $q = (q_1, q_2, \ldots, q_n)$, $q_j \geq 0$, $\sum_{j=1}^{m} q_j = 1, \quad j = \overline{1,m}$.

Функція середнього виграшу першого гравця має вигляд

$$H(p,q) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} p_i q_j = p \cdot P \cdot q^T . \tag{7.12}$$

Як і у випадку чистої стратегії, перший гравець прагнутиме максимально збільшити свій середній виграш, другий - мінімізувати його.

Нижньою та верхньою цінами матричної гри в змішаних стратегіях будуть:

$$\alpha = \max_p \min_q H(p,q), \; \beta = \min_q \max_p H(p,q). \tag{7.13}$$

За аналогією з іграми, що мають сідлові точки в чистих стратегіях, оптимальними змішаними стратегіями гравців є набори p_0, q_0, які задовольняють рівності:

$$\alpha=\beta=H(p_0,q_0) \tag{7.14}$$

Величина $v=H(p_0,q_0)$ називається ціною гри.

Еквівалентним визначенням оптимальних змішаних стратегій (див. теорему 7.1) є існування сідлової точки (p_0,q_0), тобто виконання нерівностей:

$$H(p,q_0) \leq H(p_0,q_0) \leq H(p_0,q). \tag{7.15}$$

7.4 Оптимальні стратегії та їхні властивості

При використанні змішаних стратегій також виникає питання, чи завжди матрична гра має рішення (оптимальні стратегії). На нього дає відповідь основна теорема матричних ігор фон Неймана.

Теорема 7.2. Будь-яка матрична гра має ціну (у змішаних стратегіях), а гравці мають оптимальні стратегії, тобто

$$\begin{aligned} v &= \max_p \min_q H(p,q) = \min_q \max_p H(p,q) = \\ &= \max_p H(p,q_0) = \min_q H(p_0,q) = H(p_0,q_0) \end{aligned}. \tag{7.16}$$

Доказ. Відповідно до теореми 7.1 співвідношення (7.16) еквівалентні ствердженню про існування сідлової точки у функції $H(p,q)$. Множини змішаних стратегій гравців S_p і S_q опуклі та обмежені, функція $H(p,q)$ безперервна за p, q і лінійна за цими змінними (тобто увігнуто-опукла). Відомо, що ці умови достатні для існування сідлової точки (p_0,q_0) у функції $H(p,q)$, тому для всіх наборів p,q виконується співвідношення (7.15). З нерівностей (7.15) і теореми 7.1 випливають рівності (7.16).

Оптимальні змішані стратегії мають властивості, якими можна скористатися під час їхнього пошуку.

Властивість 1. Нехай v - ціна матричної гри; тоді для того, щоб p_0 була оптимальною стратегією першого гравця, необхідно і достатньо, щоб

$$H(p_0,q) \geq v \text{ для всіх } q. \tag{7.17}$$

Аналогічно, щоб q_0 була оптимальною стратегією другого гравця, необхідно і достатньо, щоб

$$H(p,\, q_0) \leq v \text{ для всіх } p. \tag{7.18}$$

Доказ. Необхідність. Нехай p_0 - оптимальна стратегія першого гравця, тоді за теоремою 7.2 існує така змішана стратегія q_0 другого гравця, що

$$v = H(p_0,q_0) \leq H(p_0,q),$$

звідки й випливає співвідношення (7.17).

Достатність. Нехай виконується умова (7.17), тоді

$$\begin{aligned} &\min_q H(p_0,q) \geq v = \max_p \min_q H(p,q), \\ &\min_q H(p_0,q) = \max_p \min_q H(p,q) \end{aligned} \quad . \tag{7.19}$$

тобто p_0 є оптимальною стратегією.

Властивість 2. Введемо позначення

$$H(i,q) = \sum_{j=1}^{m} a_{ij} q_j, \quad i = \overline{1,n}; \tag{7.20}$$

$$H(p,j) = \sum_{i=1}^{n} a_{ij} p_i, \quad j = \overline{1,m}. \tag{7.21}$$

Вираз (7.20) представляє виграш першого гравця при використанні ним i-ої чистої стратегії, причому другий гравець використовує змішану стратегію q. Вираз (7.21) представляє виграш першого гравця, коли він використовує змішану стратегію p, а другий гравець - чисту стратегію j.

Якщо v - ціна матричної гри, то для того, щоб p_0 була оптимальною стратегією першого гравця, необхідно і достатньо виконання умови

$$H(p_0,j) \geq v, \quad j = \overline{1,m}. \tag{7.22}$$

Умова оптимальності стратегії q_0 другого гравця аналогічна:

$$H(i,q_0) \leq v, \quad i = \overline{1,n}. \tag{7.23}$$

З властивості 2 випливає: щоб установити, чи є (p,q) і v рішенням матричної гри, досить перевірити виконання нерівностей (7.22), (7.23). Невід'ємні рішення цих нерівностей, що задовольняють умовам $\sum_{i=1}^{n} p_i = 1$, $\sum_{j=1}^{m} q_j = 1$ дають рішення матричної гри.

Властивість 3. Нехай v - ціна гри, p_0, q_0 - оптимальні стратегії. Тоді, якщо $p_{0i}>0$, то $H(i,q_0)=v$; якщо $q_{0j}>0$, то $H(p_0,j)=v$.

Окрім того

$$\begin{aligned}&\min_j H(p_0,j)=\max_p \min_j H(p,j)=\\&=\min_q \max_i H(i,q)=\max_i H(i,q_0)=v\end{aligned}. \qquad (7.24)$$

Властивість 4. Нехай v - ціна гри, p_0, q_0 - оптимальні стратегії першого і другого гравців. Тоді, якщо для деякого і виконуватиметься

$$\sum_{j=1}^{m} a_{ij} q_j < v \text{, то } p_{0i}=0; \qquad (7.25)$$

якщо для деякого j виконується

$$\sum_{i=1}^{n} a_{ij} p_i > v \text{, то } q_{0j}=0. \qquad (7.26)$$

Приклад 7.3. Знайти рішення матричної гри для матриці P_1 з прикладу 7.2.

$$P_1=\begin{vmatrix}3 & 2 & 6\\ 9 & 10 & 5\end{vmatrix}$$

Використовуючи співвідношення (7.22), (7.23), одержимо систему нерівностей:

$$\begin{aligned}&3p_1+9p_2\ge v, && 3q_1+2q_2+6q_3\le v,\\&2p_1+10p_2\ge v, && 9q_1+10q_2+5q_3\le v,\\&6p_1+5p_2\ge v, && q_1+q_2+q_3=1,\\&p_1+p_2=1.\end{aligned}$$

причому p_1, p_2, q_1, q_2, $q_3 \ge 0$.

Якщо замінити нерівності в обох системах рівнянь рівностями, то в цьому випадку ліва система рівнянь виявиться неспільною. Тому припустимо $3p_1 + 9p_2 > v$ (строга нерівність), а інші нерівності замінимо на рівності. Відповідно до умови (7.26) $q_1 = 0$ і вирішуючи систему, одержимо:

$$\begin{aligned}&2p_1+10p_2=v, && 2q_2+6q_3=v,\\&6p_1+5p_2=v, && 10q_2+5q_3=v,\\&p_1+p_2=1, && q_2+q_3=1,\ (q_1=0).\end{aligned}$$

Знаходимо $p_1=5/9$, $p_2=4/9$, $q_2=1/9$, $q_3=8/9$, $v=50/9$, причому $3p_1+9p_2=51/9>v$. Рішення гри має вигляд:

$$p_0=\left(\frac{5}{9},\frac{4}{9}\right),\ q_0=\left(0,\frac{1}{9},\frac{8}{9}\right),\ v=\frac{50}{9}.$$

Отримане рішення має властивості оптимальних рішень. Так, наприклад, $p_{0i}>0$, $i=1,2$ і $H(i,q_0)=v=50/9$ (властивість 3); $H(p_0,1)=51/9>v$, $q_{01}=0$ (властивість 4).

Для вихідної матриці P прикладу 7.2 рішення гри (з урахуванням виключення рядка і стовпця) буде:

$$p_0=\left(\frac{5}{9},\frac{4}{9},0\right),\ q_0=\left(0,\frac{1}{9},\frac{8}{9},0\right),\ v=\frac{50}{9}.$$

7.5 Ігри порядку 2хm, nх2. Графічний спосіб розв'язання

Розв'язання гри з матрицею 2хm можна знайти графічним способом. Матриця має вигляд

$$P = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \end{vmatrix},$$

тобто перший гравець має дві чисті стратегії або змішану з імовірностями p і 1-p.

Розв'язання гри варто починати з відшукання сідлової точки відповідно до зауваження 1. Якщо вона існує, то визначається пара чистих стратегій для гравців. Інакше відшукуються оптимальні змішані стратегії. Відповідно до співвідношення (7.24)

$$\begin{aligned} v &= \max_{0 \le p \le 1} \min_{1 \le j \le m} [a_{1j} \cdot p + a_{2j} \cdot (1-p)] = \\ &= \max_{0 \le p \le 1} \min_{1 \le j \le m} [(a_{1j} - a_{2j}) p + a_{2j}] \end{aligned}. \qquad (7.27)$$

Відповідно до виразу (7.27) для графічного розв'язання гри на відрізку [0,1] треба побудувати m лінійних функцій $L=(a_{1j}-a_{2j})p+a_{2j}$, j=1,$m$, знайти їх нижню огинаючу, що відповідає операції $\min\limits_{1 \le j \le m}$, і на цій огинаючий знайти точку максимуму М. Абсциса точки визначає стратегію p_0, ордината - ціну гри v. Оскільки нижня огинаюча є опуклою кусочно-лінійною функцією, то точка максимуму лежить на перетині двох прямих, або через неї проходить пряма, що паралельна осі абсцис (рис. 7.1).

В першому випадку другий гравець використовує дві чисті стратегії 2 і 3 (тобто q_1 = 0) в оптимальній стратегії, у другому - одну (2).

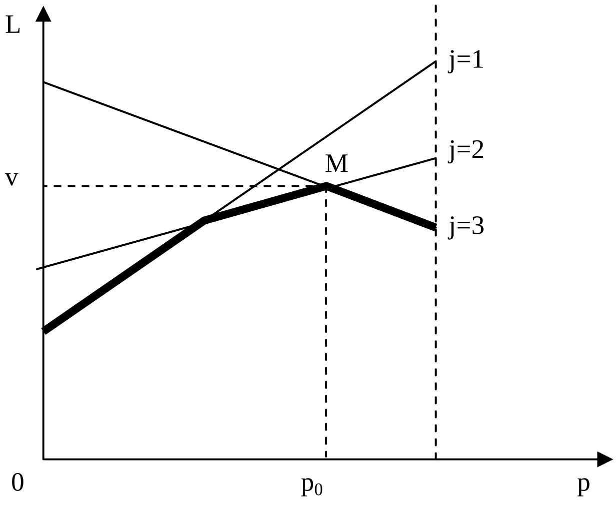


Рисунок 7.1 – Пошук максимуму

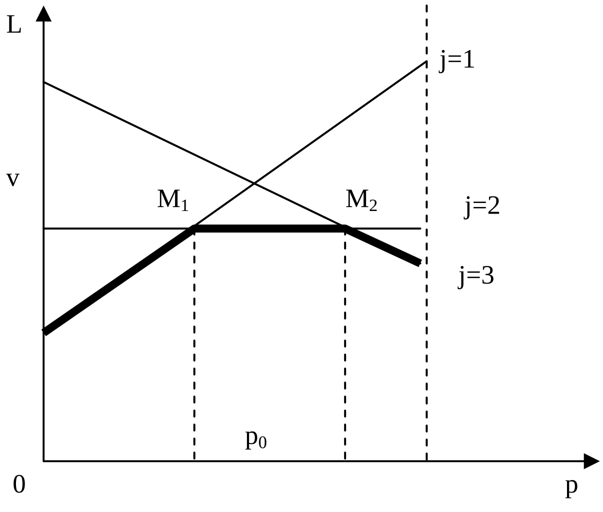


Рисунок 7.2 – Альтернативний оптимум

Точні значення p_0, q_{0j} можна одержати, вирішуючи системи рівнянь для визначення координат точки М. Для графіка на рисунку 7.1 визначимо p_0 з рівнянь (j = 2,3):

$$\begin{aligned} (a_{12} - a_{22}) \cdot p + a_{22} &= v, \\ (a_{13} - a_{23}) \cdot p + a_{23} &= v. \end{aligned} \qquad (7.28)$$

Значення q_2, q_3 (q_1=0) знаходимо відповідно до властивості 3 з рівнянь:

$$a_{12}q_2+a_{13}q_3=v,$$
$$a_{22}q_2+a_{23}q_3=v, \quad q_2+q_3=1, \tag{7.29}$$

або

$$(a_{12} - a_{22})q_2 + (a_{13} - a_{23})q_3 = 0, \tag{7.30}$$
$$q_2+q_3=1.$$

Для графіка на рисунку 7.2 перший гравець має нескінченно багато оптимальних стратегій p_0, 1-p_0. В другого гравця є чиста оптимальна стратегія j_0=2, причому ціна гри $v=a_{12}=a_{22}$ (випливає з рівняння (7.29) при q_2 = 1, q_3 = 0).

Для часткового випадку - гри 2x2 (m=2), вирішуючи системи рівнянь (7.28), (7.29) при j=1,2, можна знайти оптимальні стратегії:

$$p_{01} = p_0 = \frac{a_{22}-a_{21}}{a_{11}+a_{22}-a_{12}-a_{21}} = \frac{a_{22}-a_{21}}{A_0};\ p_{02} = 1-p_0 = \frac{a_{11}-a_{12}}{A_0} = \frac{a_{22}-a_{21}}{A_0}; \tag{7.31}$$

$$q_{01} = \frac{a_{22}-a_{12}}{A_0}; \quad q_{02} = 1-q_{01} = \frac{a_{11}-a_{21}}{A_0}; \tag{7.32}$$

$$v = \frac{a_{22}a_{11}-a_{12}a_{21}}{A_0}; \tag{7.33}$$

де $A_0=a_{11}+a_{22}-a_{12}-a_{21}$.

При грі nx2 матриця виграшів має вигляд

$$P = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \cdots & \cdots \\ a_{n1} & a_{n2} \end{vmatrix},$$

$$v = \min_{0\le q\le 1} \max_{1\le i\le n}[(a_{i1}-a_{i2})q+a_{i1}]. \tag{7.34}$$

Аналогічно попередньому на відрізку [0,1] будують n лінійних функцій $(a_{i1}-a_{i2})q+a_{i2}$, i = 1,n, далі знаходять їхню верхню огинаючу і на ній - точку мінімуму М (рис. 7.3.).

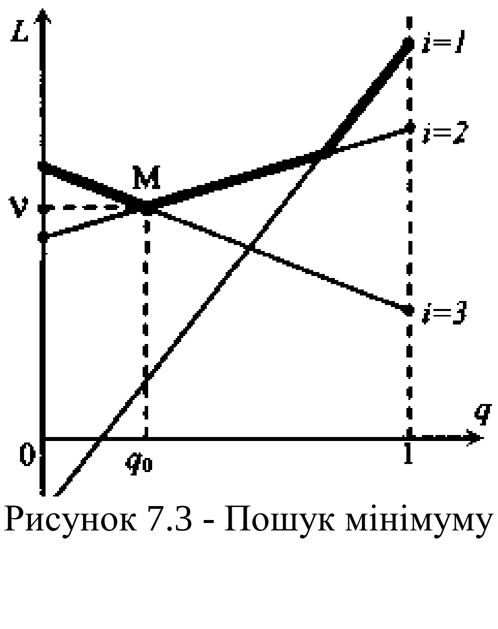


Рисунок 7.3 - Пошук мінімуму

Величина q_0 визначається з рівнянь (i = 1,2):

$$(a_{11}-a_{12})q + a_{12}=v,$$
$$(a_{21}-a_{22})q + a_{22}=v, \qquad (7.35)$$

тобто $q_{01}=q_0$, $q_{02}=1-q_0$.

Значення p_1, p_2 (p_3=0) визначаються за співвідношеннями

$$a_{11}p_1 + a_{21}p_2=v,$$
$$a_{12}p_1 + a_{22}p_2=v,\ p_1 + p_2=1. \qquad (7.36)$$

Зауваження 2. При розв’язанні ігор 2хm, nх2 гравці, що користуються m (або n) чистими стратегіями, застосовують змішану стратегію з комбінації двох чистих стратегій. Ця властивість є справедливою і для довільної гри nхm, що використовується в алгоритмах пошуку оптимальних стратегій (метод Шеплі-Сноу).

Приклад 7.4. Знайти графічним методом розв’язок гри з матрицею P_1 з прикладу 7.3. Маємо матрицю розмірами 2х3:

$$P_1 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{vmatrix} = \begin{pmatrix} 3 & 2 & 6 \\ 9 & 10 & 5 \end{pmatrix}.$$

Відповідно до рівностей (7.27) будуємо графік з прямими (рис. 7.4):

L_1(j =1): $(a_{11}-a_{21})p + a_{21}=-6p + 9$,

L_2(j =2): $(a_{12}-a_{22})p + a_{22}=-8p +10$,

L_3(j=3): $(a_{13}-a_{23})p + a_{23}=p + 5$.

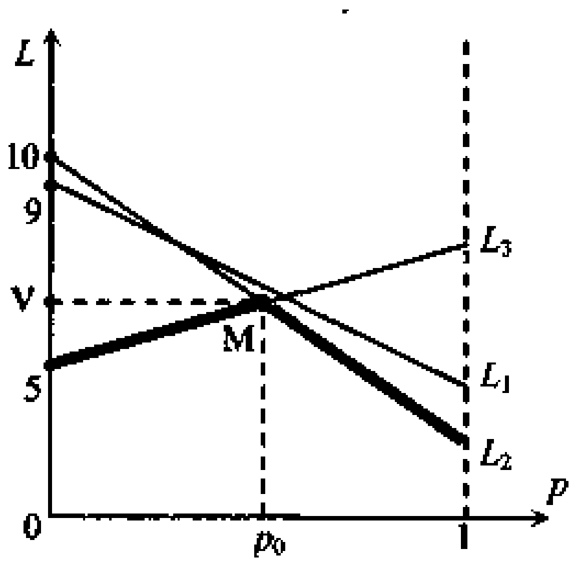


Рисунок 7.4 – Розв’язання задачі

Координату p_0 точки перетинання прямих L_2 і L_3 знайдемо з рівняння

$-8p + 10 = p + 5$, $p = p_0=p_{01}$=5/9, p_{02}=1- 5/9=4/9, v=5/9 + 5 =50/9.

Використовуючи властивість 3, маємо ($q_1 = 0$):

$$2q_2 + 6q_3 =v,\quad 10q_2+5q_3=v,\ q_2+q_3=1,$$

звідки $25q_3 =4v$, q_3 =8/9, q_2 =1/9.

7.6 Зведення матричної гри до завдання лінійного програмування

В основі відомих методів розв'язання ігор nхm (метод Шеплі-Сноу, метод послідовного наближення Брауна) лежить перебір, що зумовлює великий обсяг обчислень із зростанням значень n, m. Часто матрична гра вирішується простіше як завдання лінійної оптимізації (програмування).

Передбачається, що ціна гри $v>0$. Цього можна домогтися, зробивши елементи матриці $a_{ij}>0$ шляхом додавання до них досить великого числа - при цьому оптимальні стратегії гравців не змінюються. Нехай матриця $P=(a_{ij})$ має розмір nхm. Відповідно до нерівностей (7.22), (7.23) оптимальні змішані стратегії гравців і ціна гри v мають задовольняти співвідношенням:

$$\sum_{i=1}^{n} a_{ij} p_i \geq v, \quad j=\overline{1,m}; \ \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad i=\overline{1,n}; \tag{7.37}$$

$$\sum_{j=1}^{m} a_{ij} q_j \leq v, \quad i=\overline{1,n}; \ \sum_{j=1}^{m} q_j = 1, \quad q_j \geq 0, \quad j=\overline{1,m}. \tag{7.38}$$

Розділивши нерівності на $v>0$ і ввівши позначення $x_i=p_i/v$, $y_j=q_j/v$, одержимо

$$\sum_{i=1}^{n} a_{ij} x_i \geq 1; \ \sum_{i=1}^{n} x_i = \frac{1}{v}, \quad x_i \geq 0, \quad i=\overline{1,n}; \tag{7.39}$$

$$\sum_{j=1}^{m} a_{ij} y_j \leq 1; \ \sum_{j=1}^{m} y_j = \frac{1}{v}, \quad y_j \geq 0, \quad j=\overline{1,m}. \tag{7.40}$$

Перший гравець прагне зробити ціну гри v максимальною, тобто розв'язання завдання (7.39) збігається до визначення таких $x_i \geq 0$, щоб

$$\sum_{i=1}^{n} x_i \to \min \quad \textit{при} \quad \sum_{i=1}^{n} a_{ij} x_i \geq 1, \quad x_i \geq 0. \tag{7.41}$$

Другий гравець прагне мінімізувати v, тобто вирішити завдання (7.40):

$$\sum_{j=1}^{m} y_j \to \max \quad \textit{при} \quad \sum_{j=1}^{m} a_{ij} y_j \leq 1, \quad y_j \geq 0. \tag{7.42}$$

Вирази (7.41) і (7.42) представляють пару двоїстих симетричних завдань. Вирішивши одну з них, що вимагає менших обчислень, можна знайти рішення іншої за допомогою теорем подвійності. Завдання мають опорні рішення - для (7.42) нульовий вектор, а для (7.41) внаслідок додатності елементів a_{ij} - вектор з одним ненульовим компонентом x_i.

Приклад 7.5. Знайти рішення гри, що задана матрицею Р прикладу 7.2

$$P = \begin{pmatrix} 3 & 2 & 6 & 8 \\ 9 & 10 & 5 & 6 \\ 7 & 8 & 4 & 5 \end{pmatrix}.$$

Рівняння (7.41) містять j=4 нерівності, рівняння (7.42) i=3 нерівності; складемо систему нерівностей (7.42) і знаходимо оптимальну стратегію другого гравця.

Маємо

$$3y_1 + 2y_2 + 6y_3 + 8y_4 \le 1,\ \sum_{j=1}^{4} y_j \to \max,$$

$$9y_1 + 10y_2+5y_3+6y_4\le1, \qquad (7.43)$$

$$7y_1+8y_2+4y_3+5y_4 \le 1,\ y_j \ge 0, \quad j = \overline{1,4}.$$

Вирішуємо систему (7.43), увівши змінні y_5 - y_7:

$$3y_1 + 2y_2 + 6y_3 + 8y_4 + y_5 = 1,$$

$$9y_1 + 10y_2+5y_3+6y_4+y_6=1, \qquad (7.44)$$

$$7y_1+8y_2+4y_3+5y_4 +y_7= 1,\ y_j \ge 0, \quad j = \overline{1,7}.$$

$$z = \sum_{j=1}^{4} y_j \to \max, \text{ або } \bar{z} = -\sum_{j=1}^{4} y_j = -y_1 - y_2 - y_3 - y_4 \to \min.$$

За базисні змінні приймаємо y_5, y_6, y_7; тоді опорний план $y^{(0)}$ = (0,0,0,0,1,1,1) і рівняння (7.44) приймають вигляд

$$y_5+3y_1+2y_2+6y_3+8y_4=1,$$

$$y_6 + 9y_1 + 10y_2 + 5y_3 + 6y_4 = 1, \qquad (7.45)$$

$$y_7 +7y_1+8y_2+4y_3 +5y_4=1.$$

Цільова функція $\bar{z}$ містить вільні змінні з коефіцієнтом -1, тому вводимо в базис y_2. Знаходимо $\theta_i = \frac{1}{2}, \frac{1}{10}, \frac{1}{8}$; з них мінімальне $\frac{1}{10}$, тому виводимо з базису y_6. Одержуємо систему рівнянь

$$y_5 + \frac{6}{5}y_1 + 5y_3 + \frac{34}{5}y_4 - \frac{1}{5}y_6 = \frac{4}{5},$$

$$y_2 + \frac{9}{10}y_1 + \frac{1}{2}y_3 + \frac{6}{10}y_4 + \frac{1}{10}y_6 = \frac{1}{10}, \qquad (7.46)$$

$$y_7 - \frac{1}{5}y_1 + \frac{1}{5}y_4 - \frac{4}{5}y_6 = \frac{1}{5}.$$

Підставляючи y_2 у цільову функцію $\bar{z}$, маємо

$$\bar{z} = -\frac{1}{10} - \frac{1}{10}y_1 - \frac{1}{2}y_3 - \frac{4}{10}y_4 + \frac{1}{10}y_6,$$

мінімальне значення має коефіцієнт при y_3 (-1/2), тому вводимо в базис y_3. Значення $\theta_i = \frac{4}{25}, \frac{1}{5}$, найменше $\theta_i = \frac{4}{25}$, тому виводимо з базису y_5. Одержуємо

$$y_3 + \frac{6}{25}y_1 + \frac{34}{25}y_4 + \frac{1}{5}y_5 - \frac{1}{25}y_6 = \frac{4}{25},$$

$$y_2 + \frac{39}{50}y_1 - \frac{4}{50}y_4 - \frac{1}{10}y_5 + \frac{6}{50}y_6 = \frac{1}{50}, \qquad (7.47)$$

$$y_7 - \frac{1}{5}y_1 + \frac{1}{5}y_4 - \frac{4}{5}y_6 = \frac{1}{5},$$

$$\bar{z} = -\frac{9}{50} + \frac{1}{50}y_1 + \frac{14}{50}y_4 + \frac{1}{10}y_5 + \frac{4}{50}y_6. \qquad (7.48)$$

Коефіцієнти при вільних змінних функції $\bar{z}$ додатні, тому план $y^{(2)} = (0, \frac{1}{50}, \frac{4}{25}, 0, 0, 0, \frac{1}{50})$ є рішенням завдання.

Оскільки $\min \bar{z} = -\frac{9}{50}$, то $\max z = \frac{9}{50}$ і ціна гри v = 50/9.

З урахуванням співвідношення $q_i = y_i v$ визначаємо оптимальну стратегію другого гравця $q_0 = (0, \frac{1}{9}, \frac{8}{9}, 0)$.

Для симетричного двоїстого завдання компоненти оптимального плану $x_{опт}$ дорівнюють значенням коефіцієнтів цільової функції $\bar{z}$: $x_j = \gamma_{m+j} = \gamma_{4+j}$. З співвідношення (7.48) одержуємо

$$x_1 = \gamma_5 = 1/10,\ x_2 = \gamma_6 = 4/50,\ x_3 = \gamma_7 = 0.$$

і з урахуванням рівності $p_i = x_i \cdot v$ знаходимо оптимальну стратегію p_0 першого гравця: $p_0 = (\frac{5}{9}, \frac{4}{9}, 0)$.

Контрольні запитання

1. Сформулюйте предмет теорії ігор як наукової дисципліни.
2. Який зміст вкладається в поняття «гра»?
3. Для опису яких економічних ситуацій можна застосовувати апарат теорії ігор?
4. Яку гру називають антагоністичною?
5. Чим однозначно визначаються матричні ігри?
6. Поясніть, у чому полягають принципи максиміна та мінімакса?
7. За яких умов можна говорити, що гра має сідлову точку?
8. Наведіть приклади ігор, які мають сідлову точку та у яких вона відсутня.
9. Які підходи існують до визначення оптимальних стратегій?
10. Поясніть, що називають «ціною гри»?
11. Дайте визначення поняттю «змішана стратегія».
12. Сформулюйте основну теорему матричних ігор.

ТЕМА 8
БАГАТОКРИТЕРІАЛЬНА ОПТИМІЗАЦІЯ

8.1 Постановка проблеми багатокритеріальної оптимізації

Під дослідженням операцій розуміють застосування математичних кількісних методів для обґрунтування рішень у всіх галузях цілеспрямованої людської діяльності. Основними етапами рішення будь-якої задачі в дослідженні операцій є:

1) побудова моделі;
2) вибір критерію оптимальності;
3) знаходження оптимального рішення.

Для підходу дослідження операцій характерні наступні риси:

- використовувані моделі мають об'єктивний характер. Побудову моделей розглядають в рамках дослідження операцій як засіб відбиття об'єктивно існуючої реальності. Коли модель, що правильно відбиває дійсність, знайдено, критерій оптимальності встановлений, оптимальне рішення може бути отримане **єдино можливим** чином. Інакше кажучи, спираючись на ті самі дані, різні фахівці-аналітики мають одержувати однакові результати. Ця вимога означає, що діяльність людей, описувана моделлю, підлегла вимогам **доцільності**;

- керівник одержує науково обґрунтоване рішення. За замовленням керівника аналітик досліджує організацію, зовнішнє середовище та намагається побудувати адекватну модель. У цій роботі сама ОПР найчастіше не потрібна. Можна сказати, що керівник дає замовлення та одержує готове рішення. Все інше роблять аналітики-фахівці з дослідження операцій;

- існує об'єктивний критерій успіхів у застосуванні методів дослідження операцій. Якщо проблема, що вимагає рішення, ясна і критерій визначений, то аналітичний метод відразу показує, наскільки нове рішення краще за старе. Оптимальне рішення проблеми безглуздо заперечувати.

Аналіз багатьох реальних практичних проблем, з якими зіштовхувалися фахівці з дослідження операцій, природно призвів до появи класу **багатокритеріальних завдань**. У загальному випадку постановки завдань багатокритеріальної оптимізації є коректнішими за постановки завдань математичного програмування. Це пов'язано з тим, що будь-яка операція є сукупністю цілеспрямованих дій і проведення практично будь-якої операції, як правило, припускає досягнення не однієї, а кількох цілей.

Для багатокритеріальних задач характерним є те, що модель, яка описує множину припустимих рішень, є об'єктивною, але якість рішення оцінюється за багатьма критеріями. Для вибору найкращого варіанту рішення необхідний компроміс між оцінками за різними критеріями. В умовах задачі відсутня інформація, що дозволяє знайти такий компроміс. Отже, його не можна визначити на підставі об'єктивних розрахунків.

За наявності багатьох критеріїв задачі вибору найкращого рішення мають наступні особливості. Задача має унікальний характер - немає статистичних даних, що дозволяють обґрунтувати співвідношення між різними критеріями. На момент ухвалення рішення принципово відсутня інформація, що дозволяє об'єктивно оцінити можливі наслідки вибору того або іншого варіанту рішення. Але оскільки рішення так чи інакше треба прийняти, то нестачу інформації необхідно заповнити. Це можуть зробити лише люди на підставі їхнього досвіду та інтуїції.

У загальному випадку математичне формулювання задачі багатокритеріальної оптимізації з множиною припустимих розв’язків D і векторною цільовою функцією

$$f(X)=(f_1(X),\ldots,f_m(X))^T$$

можна записати так:

$$f_k(X) \to \underset{X\in D}{extr}, \quad k=\overline{1,m}\,.$$

Будь-який скалярний критерій вигляду

$$f_j(X) \to \max_{X \in D}$$

можна замінити еквівалентним скалярним критерієм

$$-f_j(X) \to \min_{X \in D}.$$

Отже задачу багатокритеріальної оптимізації можна сформулювати таким чином:

$$f_k(X) \to \min_{X \in D}, \quad k = \overline{1,m}, \tag{8.1}$$

або у векторній формі:

$$f(X) \to \min_{X \in D}. \tag{8.2}$$

Задачу дослідження операцій називають некоректною, якщо вона не має розв'язку. На початковому етапі розвитку дослідження операцій некоректними вважали задачі багатокритеріальної оптимізації, а як обґрунтування приводилися наступні міркування.

Нехай для кожного $k = \overline{1,m}$ елемент X_k множини D є розв'язком задачі математичного програмування

$$f(X) \to \min_{X \in D}.$$

В цьому випадку, згідно (8.1), якщо

$$X_k \equiv X_0, \quad k = \overline{1,m},$$

то X_0 - розв'язок задачі багатокритеріальної оптимізації (8.2). Але, як правило $X_k \neq X_j, \quad k \neq j$. Тому в загальному випадку треба очікувати, що задача багатокритеріальної оптимізації (8.2) не має розв'язку.

Приклад 8.1. Розглянемо просту задачу багатокритеріальної оптимізації з множиною припустимих розв'язків

$$D = \left\{ (x_1 x_2)^T : (x_1 - 2)^2 + (x_2 - 2)^2 \le 1 \right\}$$

та критеріями оптимальності

$$f_k(X) = x_k \to \min_{X \in D}, \quad k = \overline{1,2}.$$

З геометричних міркувань видно (рис. 8.1), що

$$X_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Оскільки $X_1 \neq X_2$, початкова задача не має розв'язку.

Отже, одночасне досягнення мінімуму за всіма скалярними критеріями $f_k(X)$ на одному розв'язку X в загальному випадку неможливе. Вихід полягає в пошуку певного компромісу в досягненні локальної мети. ОПР має сформулювати певний принцип компромісу і дотримуватися його під час вибору оптимального розв'язку. Принцип компромісу має визначити властивості оптимального розв'язку і дати відповідь на головне запитання: у якому смислі оптимальний розв'язок кращий за всі інші розв'язки? Окрім цього, цей оптимальний розв'язок має належати множині припустимих розв'язків задачі. Кількість можливих принципів компромісу дуже велика. Тому під час розв'язання багатокритеріальних задач виникає низка проблем, що мають не обчислювальний, а концептуальний характер.

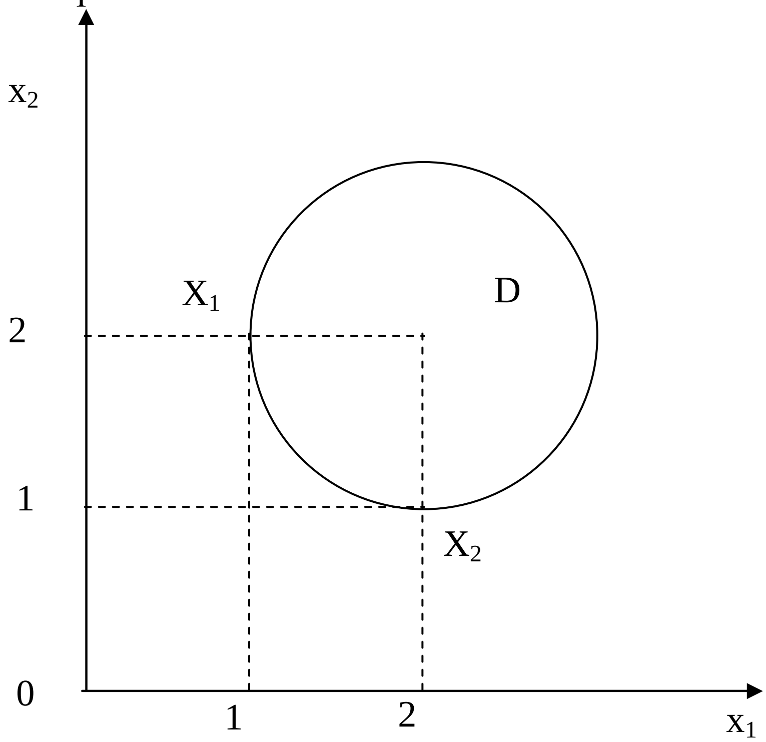


Рисунок 8.1 – Ілюстрація прикладу 8.1

На перший погляд простим виходом з ситуації, що склалася, є зведення некоректної задачі багатокритеріальної оптимізації (8.2) до відповідної задачі математичного програмування шляхом виділення з множини скалярних цільових функцій однієї основної і використання інших для формування додаткових обмежень, що накладають на множину D припустимих розв’язків (тобто шляхом виділення з множини скалярних критеріїв одного основного критерію оптимальності і переведення решти критеріїв в розряд обмежень).

Трудомісткість та проблематичність коректної реалізації розглянутого підходу в загальному випадку пов'язана як з труднощами вибору однієї основної цільової функції $f_i(X)$ з множини скалярних цільових функцій, так і з обґрунтованим призначенням верхніх меж для скалярних критеріїв, що переводяться в обмеження.

8.2 Основні множини ефективних рішень (альтернатив)

Розглянемо один з підходів до розв’язання задач багатокритеріальної оптимізації. Нехай припустимий розв’язок X_1 задачі багатокритеріальної оптимізації (8.2) є переважнішим за припустимий розв’язок X_2, тобто $X_1 \succ X_2$, якщо $f(X_1) < f(X_2)$. Остання нерівність означає, що $f_k(X_1) \le f_k(X_2)$, $k=\overline{1,m}$, причому серед цих m нерівностей є хоч би одна строга. Припустимі розв’язки X_1, X_2 задачі багатокритеріальної оптимізації (8.2) називатимемо еквівалентними розв’язками $X_1 \sim X_2$, якщо $f(X_1)=f(X_2)$. Відзначимо, що:

а) у загальному випадку з $X_1 \sim X_2$, не випливає $X_1 = X_2$;

б) якщо припустимий розв'язок X_1 є строго переважнішим за припустимий розв'язок X_2, а припустимий розв'язок X_2 є строго переважнішим за припустимий розв'язок X_3, то припустимий розв'язок X_1 є строго переважнішим за, припустимий розв'язок X_3, тобто

$$\left((X_1 \succ X_2) \wedge (X_2 \succ X_3)\right) \Rightarrow (X_1 \succ X_3);$$

в) якщо X_1 та X_2 - еквівалентні розв'язки, X_2 та X_3 – так само еквівалентні розв'язки, то X_1 та X_3 будуть еквівалентними розв'язками, тобто

$$\left((X_1 \sim X_2) \wedge (X_2 \sim X_3)\right) \Rightarrow (X_1 \sim X_3).$$

Повна множина альтернатив $\Omega = f(D)$ - множина можливих значень векторної цільової функції f у задачі багатокритеріальної оптимізації (8.2), породжувана множиною D припустимих розв'язків. Нехай m=2, тобто цільова функція $f(X)=(f_1(X), f_2(X))$, довільно обираємо припустимий розв'язок X_0, якому відповідає значення цільової функції $f(X_0)=(f_1(X_0), f_2(X_0))^T$ та підмножина $\Omega^1 \in \Omega$, породжувана підмножиною D^1 припустимих розв'язків. Отже, будь-який припустимий розв'язок з множини D^1 є строго переважнішим за припустимий розв'язок $X_0 \in D$, тобто

$$X_1 \succ X_0, \quad X_1 \in D_1 \subset D.$$

Продовжуючи ці міркування приходимо до висновку, що з повної множини альтернатив Ω можна виділити підмножину Ω^*, яка має дуже цінні властивості:

а) для припустимих розв'язків з множини Ω^* у множині D припустимих розв'язків задачі багатокритеріальної оптимізації (8.2) вже немає строго переважніших припустимих розв'язків;

б) припустимі розв'язки з множини D^* або еквівалентні, або незіставні в смислі строгої переваги.

Множину Ω^* називають множиною Парето або множиною компромісу. На множині Парето зменшення значень одного з скалярних критеріїв можна досягнути лише ціною збільшення значень іншого. Це спостерігається і в загальному випадку, оскільки будь-які припустимі розв'язки X_1, X_2, або еквівалентні, і для них $f(X_1)= f(X_2)$, або незіставні, тобто для пар координат $f_k(X_1)$, $f_k(X_2)$ векторів $f(X_1)$, $f(X_2)$ мають місце хоч би два з наступних трьох співвідношень: $f_i(X_1)>f_i(X_2)$, $f_j(X_1)=f_j(X_2)$, $f_n(X_1)<f_n(X_2)$.

Підмножина D^* множини D припустимих розв'язків є узагальненим розв'язком задачі багатокритеріальної оптимізації (8.2). Цей узагальнений розв'язок в загальному випадку може складатися більш ніж з одного елементу, і тоді ОПР стикається з проблемою вибору одного припустимого розв'язку з множини еквівалентних і незіставних рішень.

8.3 Підходи до розв'язання задачі багатокритеріальної оптимізації

Розв'язки, що утворюють множину Парето, характерні тим, що для жодного з них не існує домінуючого розв'язку. Множину таких точок називають множиною точок оптимальних за Парето або називають областю компромісів (переговорною множиною). Множину векторних оцінок, що відповідають множині неефективних точок (домінуємих розв'язків), називають областю згоди $\Omega_з$. У області $\Omega_з$ немає суперечності між частковими критеріями оптимальності, оскільки кожну точку $X \in D$ можна змінити таким чином, що одночасно будуть поліпшені всі часткові критерії. Якщо область критеріїв Ω складається тільки з області згоди $\Omega_з$, то існує єдина точка $X_{opt} \in D$, в якій всі часткові критерії узгоджені один з одним в тому смислі, що під час руху до точки X_{opt} усі $f_i(X)$ одночасно поліпшуються. Усі часткові критерії досягають мінімуму в точці X_{opt}.

Проте така ситуація зустрічається украй рідко. Найбільш типовим є випадок, коли часткові критерії є суперечливими і мінімум за кожним з них досягається в різних точках. В цьому випадку зменшення одного часткового критерію призводить до збільшення інших часткових критеріїв.

Виділення множини Парето багатокритеріальної задачі оптимізації часто не є задовільним рішенням. Це пов'язано з тим, що при достатньо великій початковій множині варіантів множина Парето виявляється неприпустимо великою для того, щоб ОПР була б в змозі здійснити вибір самостійно. Отже, виділення множини Парето можна розглядати лише як попередній етап оптимізації, і виникає проблема подальшого скорочення цієї множини.

Для вибору однієї оптимальної стратегії з множини ефективних рішень в кожній конкретній багатокритеріальній задачі необхідно використовувати додаткову інформацію про мету операції, тобто ту інформацію, яка під час завдання векторного критерію залишилася неформалізованою і тому невикористаною.

Розглянемо деякі найпростіші способи скорочення Парето-оптимальної множини.

Лексикографічна оптимізація (ранжирування критеріїв). Додаткова інформація, що допомагає у виборі розв'язку, може полягати в тому, що скалярні цільові функції $f_k(X)$ у задачі векторної оптимізації (8.1) впорядковані відповідно до їх значущості. В цьому випадку номер цільової функції відображає ранг (пріоритет) відповідного скалярного критерію.

Нехай Ω^* - множина Парето для задачі векторної оптимізації (8.1) і номер скалярної цільової функції $f_k(X)$ відповідає її рангу. Процедуру вибору єдиного рішення з підмножини D^* почнемо з використання критерію першого рангу. Вважаємо

$$q_1 = \min_{X \in D^*} f_1(X),$$

тобто D_1^* містить всі припустимі розв'язки з D^*, які мінімізують в D^* цільову функцію першого рангу. Далі переходимо до цільової функції другого рангу і вважаємо

$$q_2 = \min_{X \in D^*_1} f_2(X),$$

тобто множина D_2^* містить всі припустимі розв'язки з D_1^*, які мінімізують цільову функцію другого рангу і так далі. Переходимо до цільової функції (m-1)-го рангу і вважаємо

$$q_{m-1} = \min_{X \in D^*_{m-2}} f_{m-1}(X),$$

тобто множина D_{m-1}^* містить всі припустимі розв'язки з D_{m-2}^*, які мінімізують критерій (m-1)-го рангу. Оскільки $D_{m-1}^* \subset D_{m-2}^* \subset \ldots \subset D_1^* \subset D^*$, то для завершення процедури розв'язання задачі (8.1) в умовах ранжирування критеріїв залишилося розв'язати задачу математичного програмування

$$f_m(X) \to \min_{X \in D^*_{m-1}}.$$

Для реалізації запропонованої процедури розв'язання задачі багатокритеріальної оптимізації підмножині D_{m-1}^* множини D має відповідати підмножина множини Парето, що складається більш ніж з одного елементу. Оскільки в загальному випадку ця умова може не виконуватися, то на практиці для розв'язання задачі багатокритеріальної оптимізації частіше використовують метод, відомий як метод компромісів.

Припустимо, що для скалярної цільової функції $f_k(X)$ призначено припустиму поступку $\delta_k>0$, яка визначає величину припустимого відхилення значення критерію k-го рангу від його мінімального значення

$$\rho_k = \min_{X \in D^*} f_k(X)$$

на множині D^*. Очевидно, що для кожного $k = \overline{1, m-1}$ поступка δ_k визначає певну підмножину в множині D^*. Якщо $m>2$, ця підмножина може виявитися порожньою. В цьому випадку поступки обрані невдало і необхідна їх корекція.

Відзначимо, що практична реалізація розглянутого методу пов'язана, принаймні з двома принциповими труднощами:

а) необхідністю ранжирування скалярних критеріїв;

б) призначенням поступок та їх корекцією.

Обидві операції формально не визначаються і виконуються експертним шляхом. Звернення до експертів неминуче під час розв'язання задач багатокритеріальної оптимізації, оскільки необхідна додаткова інформація, що дозволяє ввести відношення переваги на підмножині D^*, яка відповідає множині Парето $\Omega^* \subset \Omega$. Проте, процедури ранжирування скалярних критеріїв та визначення поступок за ними не завжди прості для експертів. Тому застосування розглянутого підходу, як правило, обмежують лише тими ситуаціями, в яких експерти можуть кваліфіковано подолати обидві відзначені труднощі.

Принцип справедливої абсолютної поступки. Згідно цьому принципу, справедливим є такий компроміс, за якого сумарний абсолютний рівень підвищення одного або кількох скалярних критеріїв не перевершує сумарного абсолютного рівня зниження інших критеріїв.

Розглянемо дві точки А і В множини Парето. Через принцип справедливої абсолютної поступки під час переходу від А до В зміна вектору $f(X)$ характеризується величиною

$$\Delta_{абс} = \sum_{k=1}^{m} \Delta_k = \sum_{k=1}^{m} (f_k^B - f_k^A) = \sum_{k=1}^{m} f_k^B - \sum_{k=1}^{m} f_k^A,$$

де f_k^A, f_k^B - значення скалярних критеріїв в точках A і B.

Якщо $\Delta_{абс}<0$, то розв’язок, що відповідає точці B, вважають кращим порівняно з розв’язком, що відповідає точці A. Тому якнайкращим буде такий розв’язок, для якого $\Delta_{абс}\geq 0$ під час переходу в будь-яку іншу точку.

Отже, принцип справедливої абсолютної поступки збігається до мінімізації суми скалярних критеріїв на множині D:

$$\sum_{k=1}^{m} f_k(X) \to \min_{X\in D^*}.$$

Недолік цього принципу полягає в тому, що він припускає диференціацію за окремими критеріями: низьке значення суми $f_1 + f_2 + . + f_m$ може досягатися, коли одні критерії мають порівняно низький рівень, тоді як інші - порівняно високий рівень. Потенційна можливість такої диференціації є характерною для задач, в яких критерії виражені в різних одиницях вимірювання. У таких задачах критерії необхідно нормалізувати, тобто привести до єдиного, бажано безрозмірному, масштабу вимірювання.

Синтез глобального критерію. Ідея цього підходу дуже проста: для задачі багатокритеріальної оптимізації (8.1) будують глобальний скалярний критерій з цільовою функцією

$$F(X) = \Phi[f_1(X),\ldots,f_m(X)] = \Phi[f(X)], \tag{8.3}$$

що залежить від початкових скалярних цільових функцій так, щоб розв’язок задачі математичного програмування

$$F(X) \to \min_{X\in D}$$

був розв’язком початкової задачі (8.2) у смислі цього принципу компромісу.

Обмежимося лише коротким обговоренням синтезу глобального скалярного критерію (грец. synthesis - з'єднання, поєднання, складання) для задач багатокритеріальної оптимізації. Оскільки система нерівностей

$$f_k(X_0) \leq f_k(X), \quad k=\overline{1,m}, \quad X \in D$$

є еквівалентною системі нерівностей

$$l_k f_k(X_0) + c_k \leq l_k f_k(X) + c_k, \quad k=\overline{1,m}, \quad X \in D,$$

де $l_k > 0, \ k=\overline{1,m}$ і c_k можна вибрати довільно, то припустимий розв’язок X_0 є розв’язком задачі багатокритеріальної оптимізації (8.1) тоді і тільки тоді, коли X_0 є розв’язком задачі векторної оптимізації.

$$l_k f_k(X) + c_k \to \min_{X\in D}, \quad k=\overline{1,m}.$$

Виходячи з цього, розглянемо вимоги, яким має задовольняти функція $\Phi[f(X)]$ (8.3).

По-перше, функція $\Phi[f_1(X),\ldots,f_m(X)]$ має бути інваріантною щодо перетворення зрушення, тобто для будь-якого набору дійсних чисел і для будь-якого припустимого розв’язку X має виконуватися рівність

$$\Phi[f_1(X),\ldots,f_m(X)] = \Phi[f_1(X)+c_1,\ldots,f_m(X)+c_m]. \tag{8.4}$$

По-друге, функція $\Phi[f_1(X),\ldots,f_m(X)]$ має бути інваріантною щодо зміни масштабу будь-якого скалярного критерію f_k, тобто для будь-якого набору дійсних додатних чисел і для будь-якого припустимого розв'язку X має виконуватися рівність

$$\Phi[f_1(X),\ldots,f_m(X)] = \Phi[l_1 f_1(X),\ldots,l_m f_m(X)]. \tag{8.5}$$

Сформульовані вимоги означають, що який би не був набір дійсних чисел c_k, набір дійсних додатних чисел l_k і припустимий розв'язок X, вірною є рівність

$$\Phi[f_1(X),\ldots,f_m(X)] = \Phi[l_1 f_1(X)+c_1,\ldots,l_m f_m(X)+c_m]. \tag{8.6}$$

Виконання вимог (8.4), (8.5) або еквівалентної їм вимоги (9.6), що накладається на функцію $\Phi[f(X)]$ під час синтезу глобального скалярного критерію, визначуваного згідно (8.3), може забезпечуватися за різними способами. Зокрема, функція

$$F(X) = \sum_{k=1}^{m} l_k \overset{0}{f}_k(X), \tag{8.7}$$

де

$$\overset{0}{f}_k(X) = \frac{f_k(X) - \min\limits_{X\in D} f_k(X)}{\max\limits_{X\in D^*} f_k(X) - \min\limits_{X\in D^*} f_k(X)}$$

відповідає вимогам (8.4), (8.5). Окрім того, ця функція враховує і вимогу нормалізації критеріїв, оскільки замість абсолютних значень скалярних критеріїв розглядаються безрозмірні величини їх відносних відхилень від мінімальних значень. Критерій (8.7) називають нормованим скалярним критерієм.

У спеціальній літературі розглядають інший вид нормованого скалярного критерію, в якому врахована важливість окремих складових:

$$F(X) = \sum_{k=1}^{m} \lambda_k \overset{0}{f}_k(X), \tag{8.8}$$

де λ_k $\overset{0}{f}_k$ - певні параметри, для яких

$$\sum_{k=1}^{m} \lambda_k = 1, \quad 0 \le \lambda_k \le 1, \quad k = \overline{1,m}. \tag{8.9}$$

Ці параметри називають ваговими коефіцієнтами. Вагові коефіцієнти можна визначати різними способами, але будь-який підхід збігається до використання експертних оцінок. Відзначимо, що для експерта задача визначення вагових коефіцієнтів не простіше за задачу ранжирування скалярних критеріїв, оскільки в цьому випадку пріоритет кожного скалярного критерію він має виразити кількісно.

Метод «ідеальної» точки. Розглядається m-мірний простір (де - m кількість локальних критеріїв), в якому апріорі обирається вектор, що відображає «ідеальний» розв'язок (або, що теж саме, «ідеальна» точка, координатами якої є «ідеальні» значення, наприклад, мінімальні або максимальні значення локальних критеріїв). У цьому просторі вводиться певна метрика, з метою обчислення відстані між вектором розв'язку та «ідеальним». За якнайкраще обирають такий розв'язок, векторна оцінка якого найбільш близька до «ідеальної» точки. Недоліками методу є свавілля під час вибору ідеальної точки і введення метрики.

Визначимо узагальнений критерій таким чином. Покладемо $a_i = \max f_i(X)$, $i = \overline{1,m}$, тобто a_i є максимально (мінімально) можливим значенням за i-м критерієм. Покладемо $a=(a_1, a_2,\ldots, a_m)$. Точку a називають ідеальною. Смисл назви пов'язаний з тим, що такі точки оптимальні відразу за всіма критеріями, отримати більшого (меншого) значення за жодним критерієм неможливо. Як правило, точка $a \in \Omega$. Задамо для всіх точок функцію, що є евклідовою відстанню між точками $f_i(X)$ і a:

$$\rho(f_i, a) = \left[\sum_{i=1}^{m}(a_i - f_i)^2\right]^{\frac{1}{2}}.$$

За цільову функцію (узагальнений критерій) беруть вираз

$$f(X) = \sum_{i=1}^{m}\lambda_i\left[a_i - f_i(X)\right]^2,$$

де λ_i - вагові коефіцієнти.

Отже, задача оптимізації формулюється наступним чином

$$\min\sum_{i=1}^{m}\lambda_i\left[a_i - f_i(X)\right]^2, \quad X \in D.$$

Тут принцип оптимальності виражається функцією вибору, що визначається близькістю до ідеальної точки. Як ідеальну точку беруть директивні значення параметрів, задані ОПР.

Контрольні запитання

1. Що називають множиною Парето і чому її ще називають множиною компромісу?

2. У чому полягає основна ідея підходу до розв’язання задач багатокритеріальної оптимізації, пов'язаного з ранжируванням скалярних критеріїв? Які достатні умови реалізації цього підходу Ви знаєте?

3. У яких ситуаціях для розв’язання задач багатокритеріальної оптимізації використовують метод компромісів?

4. Сформулюйте вимоги, яким має задовольняти глобальний скалярний критерій задачі векторної оптимізації, і наведіть їх обґрунтування.

5. Доведіть еквівалентність вимог (9.6), (9.7) і (9.8), що пред'являються до глобального скалярного критерію задачі векторної оптимізації.

6. Викладіть загальну ідею синтезу глобального скалярного критерію для задачі векторної оптимізації.

7. Які можна запропонувати принципи оптимальності для задач багатокритеріальної оптимізації?

8. Які завдання оптимального проектування призводять до використання методу ідеальної точки?

СПИСОК ДЖЕРЕЛ

1. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. - М. : Высш. шк., 2003.
2. Гмурман В .Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. - М. : Высш. шк., 2004.
3. Ковальчук Т. В. Вища математика для економістів / Т. В. Ковальчук, В. С. Мартиненко. - Ч. 1. - К. : КНТЕУ, 2005.
4. Ковальчук Т. В. Вища математика для економістів / Т. В. Ковальчук, В. С. Мартиненко, В. І. Денисенко.-Ч. 2. - К. : КНТЕУ, 2007.
5. Кремер Н. Ш. Высшая математика для экономистов / Н. Ш. Кремер. - М.: ЮНИТИ, 2001.
6. Мартиненко В. С. Збірник задач з вищої математики / В. С. Мартиненко та ін. - Ч. 1. - К. : КНТЕУ, 2000.
7. Мартиненко В. С. Збірник задач з вищої математики / В. С. Мартиненко та ін. - Ч. 2. - К. : КНТЕУ, 2002.
8. Зайченко Ю. П. Дослідження операцій : підручник / Ю. П. Зайченко. - К. : ВІПОЛ, 2000.
9. Исследование операций в экономике : учеб. пособ. / под. ред. Н. Ш. Кремера. - М. : Банки и биржи, ЮНИТИ, 1999.
10. Таха Х. Введение в исследование операций / Х. Таха. - М. : Вильямс, 2005. – 912 с.
11. Красс М. С. Математика для экономических специальностей / М. С. Красс. - М. : Дело-М, 2002.
12. Поллард Дж. Справочник по вычислительным методам статистики / Дж. Поллард. - М. : Финансы и статистика, 1982. - 344 с.
13. Танаев В. С. Теория расписаний. Одностадийные системы / В. С. Танаев, В. С. Гордон, Я. М. Шафранский / - М. : Наука, 1984.
14. Потапов Г.Г. Экономические ресурсы повышения эффективности предприятия (Исследование операций) : Учебное пособие / Г. Г. Потапов, А. А. Илаев. - Симферополь: ИТ «АРИАЛ», 2011. – 98 с.
15. Ульянченко О. В. Дослідження операцій в економіці / О. В. Ульянченко. -Х. : Гриф, 2003.

Навчальне видання

АЧКАСОВ Анатолій Єгорович
ВОРОНКОВ Олексій Олександрович,

КОНСПЕКТ ЛЕКЦІЙ

з курсу

ВИЩА ТА ПРИКЛАДНА МАТЕМАТИКА (ДОСЛІДЖЕННЯ ОПЕРАЦІЙ 1)

(для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030601 – Менеджмент)

Відповідальний за випуск: *Т. А. Пушкар*

За авторською редакцією

Комп’ютерне верстання: *І. В. Волосожарова*

План 2015, поз. 181 Л

Підп. до друку 03. 04.2015 р. Формат 60х84/16
Друк на різографі Ум. друк. арк. 6,5
Тираж 50 пр. Зам. №

Видавець і виготовлювач:
Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@kname.edu.ua
Свідоцтво суб’єкта видавничої справи:
ДК № 4705 від 28.03.2014 р.